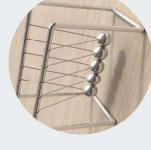
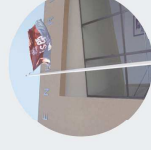




Classificação e previsão de cardiotocogramas usando algoritmos baseados em redes neuronais

DIANA FILIPA CARNEIRO BARROS

Outubro de 2012



Classificação e previsão de cardiotocogramas usando algoritmos baseados em redes neuronais

DIANA FILIPA CARNEIRO BARROS
Outubro de 2012



Classificação e previsão de cardiotocogramas usando algoritmos baseados em redes neuronais

Diana Filipa Carneiro Barros

“DISSERTAÇÃO APRESENTADA NO INSTITUTO SUPERIOR DE ENGENHARIA DO
PORTO PARA A OBTENÇÃO DO GRAU DE MESTRE EM COMPUTAÇÃO E
INSTRUMENTAÇÃO MÉDICA”

Orientador

DOUTOR JORGE MANUEL FERNANDES DOS SANTOS

4 de Outubro de 2012

Aos meus maiores exemplos de vida

Helder, Adélia e Sílvia

Agradecimentos

As primeiras palavras de agradecimento vão para o meu orientador, Professor Jorge Santos, pelo apoio, orientação e comentários sempre construtivos e válidos ao longo da realização desta dissertação.

À Dr. Maria Antónia Costa, médica no Serviço de Ginecologia e Obstetrícia do Hospital de São João, pela disponibilidade com que me recebeu e prontamente esclareceu as minhas dúvidas sobre cardiocotografia, uma área nova para mim.

Fico igualmente agradecida ao Professor João Paulo Teixeira, do Instituto Politécnico de Bragança, pelos esclarecimentos e dicas sobre séries temporais que foram úteis no decorrer do trabalho. Agradeço também ao Professor José Maria Júnior da Universidade Federal do Piauí, no Brasil, pela disponibilidade e brevidade das respostas às minhas dúvidas relativamente à aplicação de RNA na previsão de séries temporais.

Um obrigado também à empresa *Speculum*, na pessoa de Hugo Ferreira, pelas orientações fornecidas relativamente ao sistema *Omniview SisPorto*® que foram de extrema importância.

Um agradecimento especial aos meus amigos e companheiros Célio Cerqueira, Cláudia Mendes e Diana de Amorim, pela amizade, força e constante incentivo neste momento importante da minha vida e também Ana Raquel Oliveira, Cláudia Silva e Hélder Silva com quem tive a felicidade de conviver e aprender.

Aos meus pais e irmã, o meu maior orgulho e a minha verdadeira inspiração, agradeço pelo constante amor, suporte, incentivo e compreensão que tiveram ao longo destes últimos tempos para que eu pudesse cumprir mais esta meta da minha vida com sucesso. Muito obrigada pelo amor e apoio incondicional!

Serei grata a todos os que me ajudaram direta ou indiretamente na realização desta dissertação.

Resumo

A cardiocotografia (CTG), sendo um exame efetuado à grávida para avaliação do bem estar fetal, desempenha um papel importante na área da obstetrícia pois permite avaliar problemas relacionados com o feto, principalmente no período anteparto. No entanto, o resultado da análise do sinal de CTG por parte dos diferentes técnicos (obstetras) não é consensual e é aceite pela comunidade médica a existência de discrepâncias na sua avaliação.

Estas diferenças na leitura do sinal de CTG podem ser atenuadas recorrendo à introdução de sistemas automáticos de classificação que têm como objetivo principal o de garantir a aplicação dos critérios estabelecidos e o de introduzir alguma uniformização na sua avaliação.

O trabalho aqui descrito teve como objetivo principal o estudo da utilização das redes neuronais artificiais (RNA) na classificação e previsão do sinal CTG. Assim foram aplicadas metodologias de classificação baseadas em redes *feedforward* (FFNN) com variação dos parâmetros envolvidos. No trabalho aqui apresentado é também avaliada a capacidade preditiva de instantes de tempo à frente tendo por base as características do sinal e o comportamento passado que são informações importantes na determinação do futuro estado fetal. Neste estudo é testada a arquitetura *Focused Time Delay Neural Network* (FTDNN) em tarefas de previsão de um e múltiplos instantes adiante. Os resultados mostram um desempenho promissor para a classificação em diferentes comportamentos fetais no conjunto de sinais CTG utilizados bem como na previsão de instantes futuros das séries correspondentes à frequência cardíaca fetal (FCF) do CTG.

Palavras-chave: Cardiotocografia, Redes Neuronais Artificiais, Frequência Cardíaca Fetal, Classificação, Séries Temporais, Previsão

Abstract

The cardiotocography (CTG) is a clinical exam, performed in pregnant women to evaluate the fetal well-being, and has an important role in the obstetrics because it allows evaluating problems with the fetus, especially in the period before labour. However, the result of CTG analysis by different technicians (obstetricians) is not consensual and it is accepted by the medical community the presence of discrepancies in the assessment. These differences can be attenuated with the introduction of automatic classification systems whose main purpose is to ensure the application of the established criteria and to introduce uniformity in signal assessment.

The work described here aimed at studying the use of Artificial Neural Networks (ANN) in CTG classification and prediction. Different classification methodologies were applied based on feedforward neural networks (FFNN). In this work we also assess the prediction ability of ANNs based on signal characteristics and past behaviour which are important information to determine the future fetal state. In this dissertation, the *Focused Time Delay Neural Network* (FTDNN) architecture with one step and multi step ahead prediction tasks is used. The results show a good performance of the classification algorithm as well as in the prediction of the CTG signal.

Keywords: Cardiotocography, Artificial Neural Networks, Fetal Heart Rate, Classification, Time Series, Prediction

Conteúdo

Resumo	viii
Abstract	x
Índice	xii
Lista de Figuras	xiv
Lista de Tabelas	xvi
Lista de Abreviaturas	xviii
1. Introdução	1
1.1 Motivação e Objetivos	2
1.2 Estrutura da Dissertação	3
2. Conceitos Básicos	5
2.1 Cardiotocografia	5
2.1.1 Análise dos Traçados	7
2.1.2 Classificação dos Traçados	7
2.1.3 Análise automatizada do CTG	9
2.2 Redes Neurais Artificiais	13
2.2.1 Modelo do Neurónio Artificial	14
2.2.2 Funções de Ativação	15
2.2.3 Arquiteturas de rede	17
2.2.4 Processos de Aprendizagem	19
2.2.5 Medida de <i>performance</i> de classificadores	24
3. Classificação de CTG usando FFNN	31
3.1 Estado da Arte	31
3.2 Descrição do problema	32
3.3 Algoritmo desenvolvido	33
3.3.1 <i>Matlab</i>	33
3.3.2 Conjuntos de dados	34
3.3.3 Pré-Processamento dos dados	37
3.3.4 Metodologia de Classificação	39
3.4 Classificação dos conjuntos de dados	42

3.4.1	Resultados dos conjuntos de dados auxiliares	42
3.4.2	Resultados da classificação do CTG	43
3.5	Conclusões	46
4.	Previsão do sinal CTG	49
4.1	Estado da arte	49
4.2	Descrição do problema	51
4.3	Algoritmo de previsão com base na FTDNN	52
4.3.1	Séries de referência	52
4.3.2	Pré-Processamento dos dados	54
4.3.3	Criação do modelo de previsão	55
4.4	Resultados da Previsão	57
4.4.1	Resultados da previsão da série de <i>Mackey-Glass</i>	57
4.4.2	Resultados da previsão do laser	58
4.4.3	Resultados da previsão do sinal CTG	59
4.5	Conclusões	63
5.	Conclusão	67
5.1	Desenvolvimentos Futuros	69
A.	Resultados Intermédios da Classificação	71
A.1	Resultados obtidos para escolha dos parâmetros de classificação	71
A.1.1	Seleção dos algoritmos de treino	71
A.1.2	Seleção dos valores da taxa de aprendizagem e número de neurónios	71
A.2	Experiências Preliminares da classificação com o algoritmo desenvolvido	72
B.	Resultados Intermédios da Previsão	75
B.1	Divisão das séries temporais para previsão	75
B.2	Outros Resultados de Previsão obtidos	75
	Bibliografia	79

Lista de Figuras

2.1	Traçado CTG com a FCF e a UC.	6
2.2	Exemplo de um cardiotocógrafo Edan F9 Gemelar.	6
2.3	Diferentes caraterísticas identificadas no traçado CTG.	8
2.4	Visualização de sinal CTG analisado e classificado pelo Omniview-SisPorto® com os respetivos resultados quantitativos.	9
2.5	Visualização Omniview-SisPorto® : Janela do programa com apresentação simultânea de 3 simulações com os respetivos alertas.	11
2.6	Traçado CTG representativo do alerta <i>Attention</i>	13
2.7	Sinal CTG representativo do alerta <i>Very Non-Reassuring</i> com repetitivas desacelerações.	13
2.8	Exemplo de um traçado patológico com o respetivo alerta do tipo <i>Pre-Terminal</i>	14
2.9	Modelo de um neurónio artificial constituído por um conjunto de pesos, uma unidade de soma e uma função de ativação.	15
2.10	Funções de ativação mais utilizadas:(a) Função Degrau; (b) Função Logística Sigmoide com $a=1$; (c) Função Tangente Hiperbólica com $a=1$	16
2.11	Rede neuronal do tipo <i>feedforward</i> de apenas uma camada.	18
2.12	Representação de uma MLP com uma camada escondida.	18
2.13	Arquitetura genérica de uma rede FTDNN de uma camada escondida.	19
2.14	Esquema representativo de uma RNA recorrente.	19
2.15	Modelos representativos de aprendizagens: (a) Supervisionada; (b) Não supervisionada.	20
2.16	Esquema Representativo da <i>Error-correction learning</i>	21
2.17	Fluxo do sinal e retro-propagação do erro (tracejado) numa rede perceptrão simples.	22
2.18	Representação de uma matriz de confusão.	26
2.19	Curva ROC.	26
2.20	Diferença entre a Comparação de algoritmos usando a curva ROC e PR: (a) Comparação no espaço ROC; (b) Comparação no espaço PR.	27
3.1	Número de ocorrências por diferente duração do exame CTG.	37
3.2	Esquema representativo da <i>k-fold cross validation</i>	38
3.3	Janela de visualização do treino da RNA.	41
3.4	Janela de visualização da matriz de confusão para os dados Íris.	41
3.5	Matriz de Confusão obtida para uma iteração da RNA.	46

4.1	Dormidas dos turistas nas Unidades de Alojamento da Região da Madeira, no intervalo [Jan-87:Dez-09].	50
4.2	Esquema de Previsores:(a) Sem realimentação; (b) Recursivo, com alimentação.	52
4.3	Séries de Referência: (a) Série Caótica de <i>Mackey-Glass</i> ; (b) Série do Laser Caótico.	53
4.4	Série temporal referente à FCF de um traçado CTG.	54
4.5	Resultados da suavização do sinal de FCF do CTG: (a) M=25 ;(b) M=100; (c) M=200; (d) M=300.	56
4.6	Série CTG modificada com distinção dos conjuntos de treino e teste. .	57
4.7	Previsão da série de Mackey-Glass: (a) UPA ; (b) MPA.	58
4.8	Previsão da série do laser caótico: (a) UPA ; (b) MPA.	60
4.9	Previsão da série CTG UPA.	61
4.10	Previsão da série CTG 10 passos à frente.	61
4.11	Previsão da série CTG 100 passos à frente.	62
4.12	Previsão da série CTG UPA com RNA de duas camadas escondidas. .	62
4.13	Previsão da série CTG de 100 passos adiante com RNA de duas camadas escondidas.	63
4.14	Representação da série de FCF do CTG utilizada nesta experiência. .	63
4.15	Representação da série de FCF do CTG após filtragem.	64
4.16	Previsão da segunda série de CTG UPA.	65
4.17	Previsão da segunda série de CTG 10 instantes à frente admitindo duas camadas escondidas na RNA.	65
B.1	Série de <i>Mackey-Glass</i> com distinção dos conjuntos de treino e teste.	75
B.2	Série do laser com distinção dos conjuntos de treino e teste.	76
B.3	Previsão da série de CTG 5 passos à frente.	76
B.4	Resultados da suavização do segundo sinal de FCF do CTG: (a) M=5 ;(b) M=10; (c) M=25.	77
B.5	Previsão da FCF do segundo sinal de CTG 50 passos à frente.	78
B.6	Previsão das UC da segunda série CTG UPA.	78
B.7	Previsão das UC da segunda série CTG 50 passos à frente.	78

Lista de Tabelas

2.1	Descrição dos alertas do Omniview-SisPorto®.	12
2.2	Exemplos de índices de <i>performance</i>	25
3.1	Conjunto de dados reais utilizados neste trabalho.	34
3.2	Distribuição do conjunto de dados <i>Yeast</i> pelas respectivas classes.	35
3.3	Distribuição do conjunto de dados CTG pelas respectivas classes.	36
3.4	Distribuição do conjunto de dados CTG por 3 classes.	37
3.5	Diferentes tipos de normalização.	39
3.6	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados Íris usando a normalização no intervalo $[0,1]$	42
3.7	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados Íris usando a normalização no intervalo $[-1,1]$	42
3.8	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados <i>Yeast</i> usando a normalização no intervalo $[0,1]$	43
3.9	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados <i>Yeast</i> usando a normalização no intervalo $[-1,1]$	43
3.10	Resultados de <i>performance</i> da rede (MSE) com divisão de 60% para treino e 40% para teste.	44
3.11	Resultados de <i>performance</i> da rede (MSE) com divisão de 70% para treino e 30% para teste.	44
3.12	Resultados de <i>performance</i> da rede (MSE) com divisão de 90% para treino e 10% para teste.	45
3.13	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG usando a normalização no intervalo $[0,1]$	45
3.14	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG usando a normalização no intervalo $[-1,1]$	46
4.1	Erros na previsão da série de <i>Mackey-Glass</i>	59
4.2	Erros na previsão da série de laser caótico.	59
4.3	Erro na previsão da série CTG.	59
4.4	Erros na previsão da série CTG admitindo duas camadas escondidas na RNA.	61
4.5	Erros na previsão da segunda série CTG.	62
4.6	Erros na previsão da segunda série CTG com duas camadas escondidas na RNA.	64
A.1	Erros dos algoritmos de treino relativamente ao conjunto de dados Íris.	71

A.2	Resultados de <i>performance</i> da rede (MSE) para o conjunto de dados Íris.	72
A.3	Resultados de <i>performance</i> da rede (MSE) para o conjunto de dados <i>Yeast</i>	72
A.4	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 60% para treino e 40% para teste.	73
A.5	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 70% para treino e 30% para teste.	73
A.6	Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 90% para treino e 10% para teste.	73

Lista de Abreviaturas

bpm	Batimento por minuto
CTG	Cardiotocograma
ECG	Eletrocardiograma
fECG	Eletrocardiograma fetal
FCF	Frequência Cardíaca Fetal
FCM	Frequência Cardíaca Materna
FF	<i>FeedForward</i>
FFNN	<i>FeedForward Neural Network</i>
FIGO	Federação Internacional de Ginecologia e Obstetrícia
FTDNN	<i>Focused Time Delay Neural Network</i>
IA	Inteligência Artificial
MLP	Perceptrão MultiCamadas
MPA	Múltiplos Passos Adiante
MSE	<i>Mean Squared Error</i>
NARX	<i>NonLinear Autoregressive Model with Exogenous Inputs</i>
NMSE	<i>Normalized Mean-Squared Error</i>
PR	<i>Precision-Recall</i>
ROC	<i>Receiver Operating Characteristic</i>
RNA	Rede Neuronal Artificial
UC	Contrações Uterinas
UPA	Um Passo Adiante

Introdução

A cardiotocografia (CTG) é um exame complementar de diagnóstico que permite avaliar o bem-estar fetal. A monitorização eletrónica possibilita registar em simultâneo e continuamente a frequência cardíaca fetal (FCF) e a contratilidade uterina (UC) obtendo assim vários padrões que se correlacionam com os movimentos fetais. O conhecimento e análise dos mecanismos fisiológicos subjacentes a esses padrões permitem avaliar o estado clínico do feto. No entanto, a interpretação do CTG é de difícil execução devido à sua complexidade e, por conseguinte, à elevada variabilidade inter e intra observador. Este facto traduz a dificuldade da existência de uma proposta única de classificação aceite internacionalmente pois a interpretação é subjetiva e dependente da experiência individual do perito que avalia [1, 2, 3, 4, 5].

A Federação Internacional de Ginecologia e Obstetrícia (FIGO) promoveu em 1986 um encontro entre médicos especialistas de forma a chegarem a um conjunto de normas consensuais que permitissem constituir o primeiro acordo sobre aspetos essenciais do método como a terminologia, indicações técnicas e interpretação. Estas orientações permanecem atualmente como o único documento de consenso internacional sobre monitorização fetal, sendo uma referência nesta área [6].

A partir destas diretrizes, e de forma a diminuir-se esta natureza subjetiva da interpretação, cada vez mais a análise visual tem sido substituída pela análise automatizada permitindo o acesso a informação importante para o diagnóstico que pode ser ignorada na observação a olho nú. A eficácia de qualquer método de diagnóstico é por isso limitada e a análise computadorizada de CTG surge como um método para ultrapassar este problema. Os sistemas de análise de CTG computadorizados apresentam vantagens clínicas e técnicas destacando-se a reprodutibilidade, a análise sistemática do traçado, avaliação objetiva dos parâmetros, fácil acesso a informação em tempo real de vários casos e os alertas em tempo real. Contudo, estes sistemas no seu desenvolvimento encontraram dificuldades como a falta de consenso na interpretação, a instabilidade do sinal FCF no período intraparto, a necessidade de

desenvolver algoritmos de análise complexos, a constante evolução do *hardware* e *software* e as dificuldades na demonstração da sua eficácia [2].

O Omniview-SisPorto® é um exemplo de sucesso de um sistema de interpretação e avaliação do CTG. Este sistema relaciona também informação sobre o eletrocardiograma fetal (fECG), ou seja, tem em conta também mudanças do segmento ST e da onda T que podem ser importantes marcadores de *stress* fetal [2]. Trata-se de um líder em tecnologia de análise de CTG tendo motivado inúmeras publicações em jornais de referência [7, 8, 9, 10, 11].

A classificação do CTG atribui a um traçado específico um determinado grupo ou classe pré-definido com base nas suas características, determinando assim o tipo do sinal. A evolução tecnológica tem vindo a permitir o aparecimento de novos algoritmos e estratégias de classificação tendo por base métodos de inteligência artificial (IA), nomeadamente as redes neuronais artificiais (RNAs).

As RNAs são modelos que surgiram como tentativa de reprodução do cérebro humano e têm tido inúmeras aplicações em diferentes áreas. Mais concretamente, na área médica surgem em estudos de análise e classificação de imagem, reconhecimento de voz, análise de células cancerígenas, deteção de eventos epiléticos, análise de variáveis fisiológicas e avaliação de pacientes em estado crítico entre muitas outras aplicações [12, 13]. O uso destas técnicas de IA tem provado que representam uma alternativa válida a vários métodos de classificação convencional realizando operações como por exemplo aproximação de funções, análise de regressões, previsão de séries temporais ou processamento de dados [14]. No entanto é difícil implementarem-se algoritmos com regras heurísticas usadas pelos clínicos durante a realização do diagnóstico.

Para além da eficácia demonstrada na classificação de dados, tem existido um crescente interesse na utilização de RNAs para a resolução de problemas práticos no que respeita à previsão de séries temporais. Isto deve-se sobretudo à sua capacidade para capturar relações funcionais complexas subjacentes nas observações de séries temporais [15, 16, 17].

1.1 Motivação e Objetivos

O CTG é um traçado bastante útil particularmente no auxílio à determinação do momento ideal para o parto. Através da sua análise o obstetra pode antever problemas com o bebé e antecipar o parto evitando assim graves consequências. Como referido, atualmente existem vários sistemas para classificação automática de situações anómalas e que são importantes instrumentos de ajuda à decisão médica.

As RNA já foram utilizadas no campo da monitorização fetal, nomeadamente na classificação de padrões, especializadas em subtarefas específicas como por exemplo

linha base do CTG ou as desacelerações [18, 19, 20, 21].

Esta dissertação tem dois objetivos principais, sendo o primeiro o desenvolvimento de um modelo de classificação de CTG com aplicação a dados reais na tentativa de melhorar os resultados dos classificadores existentes. O ideal na aquisição de um sinal fisiológico seria conseguir, a partir dos parâmetros observados ao longo do tempo, prever acontecimentos futuros. Faz-se também nesta dissertação a tentativa de previsão de um e mais instantes de tempo do sinal de CTG.

1.2 Estrutura da Dissertação

Esta dissertação desenvolve-se ao longo de cinco capítulos. O capítulo 2 dedica-se à descrição dos conceitos teóricos mais importantes do tema de estudo. Começa por descrever o sinal CTG referindo as suas principais características, o modo como é adquirido e a importância da sua classificação. Aborda ainda o sistema automatizado de avaliação do CTG, Omniview-SisPorto®, e qual a base da respetiva classificação. Neste capítulo é feita também uma introdução às RNAs e apresentam-se as definições básicas inerentes como funções de ativação, arquiteturas de rede, métodos de aprendizagem e medidas de *performance*.

No capítulo 3 é apresentado o algoritmo desenvolvido para a classificação do CTG. Inicia-se com uma breve referência ao estado da arte nesta área, seguido da descrição dos conjuntos de dados utilizados no âmbito da classificação. Posteriormente são apresentadas as técnicas utilizadas e o modo como se desenvolveu o algoritmo e, por fim, os respetivos resultados.

A temática de previsão do sinal de CTG é apresentada no capítulo 4, onde se refere também as diferentes aplicações de RNAs para previsão de séries temporais. Para a extração de resultados são utilizadas inicialmente séries temporais caóticas, seguidas da série principal de estudo, o CTG. Em particular, este capítulo tem o objetivo de observar o desempenho de RNAs no tratamento de séries que apresentam dependência temporal. Na parte final são apresentados os principais resultados do método desenvolvido.

As principais conclusões da dissertação estão apresentadas no capítulo 5 sendo apontadas também algumas perspetivas de investigação futura.

No anexo A podem consultar-se alguns resultados intermédios obtidos pelo algoritmo de classificação criado. Por sua vez, no anexo B apresentam-se as diferentes divisões das séries temporais utilizadas na previsão com RNAs bem como experiências preliminares.

Conceitos Básicos

Este capítulo apresenta uma visão geral de conceitos teóricos que são de grande importância para a compreensão do sinal principal de estudo, bem como das técnicas aplicadas ao longo deste trabalho. Encontra-se estruturado em duas partes fundamentais: a secção 2.1 apresenta o sinal CTG, descrevendo as suas características principais, os pressupostos da classificação e o atual sistema computadorizado para a sua avaliação. A secção 2.2 descreve as características básicas das RNAs usadas na classificação e previsão como a sua arquitetura, os métodos de aprendizagem, o conceito de função de ativação, algoritmos de treino e as medidas de *performance* de classificadores.

2.1 Cardiotocografia

A deteção precoce de anomalias no feto é uma preciosa ajuda para se conseguir evitar situações perigosas que poderiam ser de difícil ou mesmo impossível resolução. A cardiotocografia é amplamente utilizada como uma ferramenta não invasiva para verificar as condições fetais em períodos ante e intraparto (períodos antes e durante o parto, respetivamente) [2, 22, 23].

Considera-se como sendo uma técnica simultânea de aquisição da FCF (com o uso de um transdutor de ultrasom *Doppler*) e das UC (medida através de um tocodinamómetro externo). No registo do CTG a FCF e as UC encontram-se expressas em batimentos por minuto (bpm) e mmHG, respetivamente [14, 24]. Na figura 2.1 pode ver-se um exemplo de um traçado de CTG onde se encontram algumas características como as acelerações e desacelerações da FCF e as UC.

Cada vez mais a monitorização eletrónica fetal é usada como um método de vigilância, quer em casos de anteparto ou intraparto. Trata-se de uma aquisição contínua de características do sinal como acelerações e desacelerações, através de equipamentos altamente eficazes e inovadores, como se pode verificar na figura 2.2,

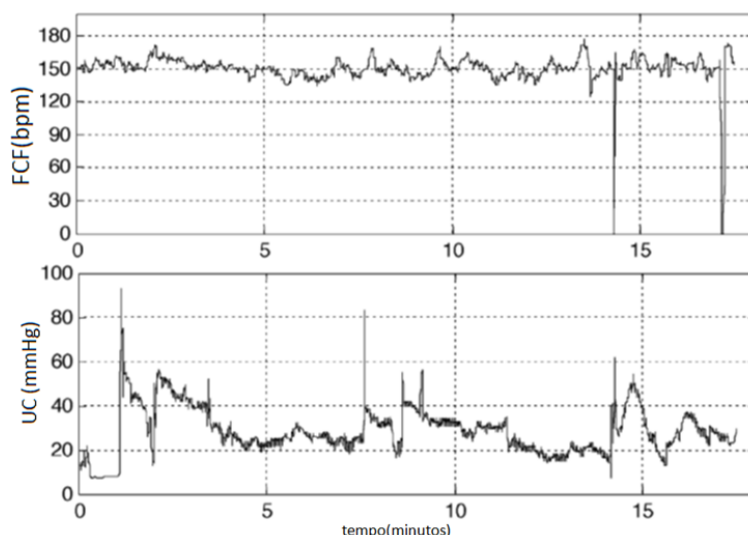


Fig. 2.1: Traçado CTG com a FCF e a UC.

o exemplo de um cardiotocógrafo.



Fig. 2.2: Exemplo de um cardiotocógrafo Edan F9 Gemelar [25].

A análise do fECG nomeadamente das modificações do segmento ST, tem vindo a aliar-se à cardiotocografia de forma a melhorar a classificação do bem estar fetal. O eletrocardiograma (ECG) corresponde a um registo de flutuações dos potenciais de ação do miocárdio durante o ciclo cardíaco, sendo o segmento ST correspondente à repolarização ventricular, isto é, ao relaxamento do coração.

O estudo deste intervalo pode identificar situações em que existe um baixo teor de oxigénio nos tecidos, designando-se este fenómeno por hipóxia. De acordo com este pressuposto, caso o segmento ST sofra uma depressão abaixo da linha base, este torna-se negativo relacionando-se portanto com uma iminente asfixia fetal [23, 26].

Mais de 60 % das mortes fetais ocorrem antes do início do parto e através da execução deste exame indolor é possível obter-se uma avaliação sobre o bem-estar fetal. Torna-se portanto perceptível a enorme importância na correta análise e posterior classificação dos traçados de CTG.

2.1.1 Análise dos Traçados

Atualmente a monitorização fetal contínua é um dos procedimentos obstétricos mais comuns no mundo e a sua correta interpretação é de extrema importância. A análise clínica do CTG encontra-se ainda muito associada à interpretação visual, isto é, largamente dependente do grau de experiência e treino do clínico envolvido. Apesar de existirem normas de consenso promovidas pela FIGO, ACOG¹, NICHHD² e RCOG³ existe ainda muita divergência, a nível mundial, relativamente à interpretação da cardiotocografia [3, 6, 24, 27, 28, 29].

As definições das características da FCF devem ser cuidadosamente analisadas para proceder-se a uma correta interpretação. A linha de base da FCF é uma dessas características e define-se como sendo o nível médio da FCF quando esta é estável na ausência de acelerações e desacelerações. Determina-se durante um período de 5 a 10 minutos e exprime-se em bpm. As acelerações caracterizam-se por um aumento transitório da FCF acima da linha de base traçada com amplitudes máximas acima de 15 bpm e duração acima dos 15 segundos. Por sua vez, as desacelerações são episódios de descida da FCF abaixo da linha base, distinguindo-se quanto à sua duração: prolongadas quando permanecem mais de 5 minutos e repetitivas se existirem 3 ou mais ocorrências em 10 minutos de aquisição ou quando ocupam mais de 50% desses segmentos do traçado. A variabilidade é outra importante referência para interpretação do CTG e, em muitos casos, não é interpretada corretamente à vista desarmada. Os intervalos entre batimentos cardíacos fetais sucessivos estão constantemente sujeitos a pequenas variações, designando-se por variabilidade a curto prazo. No entanto a diferença entre a FCF máxima e mínima no intervalo de um minuto designa-se por variabilidade a longo prazo (podendo ser normal ou reduzida) [1, 28]. Na figura 2.3 podem identificar-se algumas características no traçado de CTG.

2.1.2 Classificação dos Traçados

Considerando as características principais de um traçado CTG é necessário associá-las para determinar o tipo de situação em que se encontra o feto. Para tal existem diferentes classificações, de acordo com normas pré-estabelecidas que orientam os clínicos no diagnóstico.

Para a classificação, deve ter-se em conta não só os parâmetros relativos à FCF mas também às UC e ter sempre em atenção a relação entre estes dois sinais. As UC podem apresentar atividade normal (na fase ativa, até um máximo de 5 contrações

¹ *American College of Obstetricians and Gynecologists*

² *National Institute of Child Health and Human Development*

³ *Royal College of Obstetricians and Gynecologists*

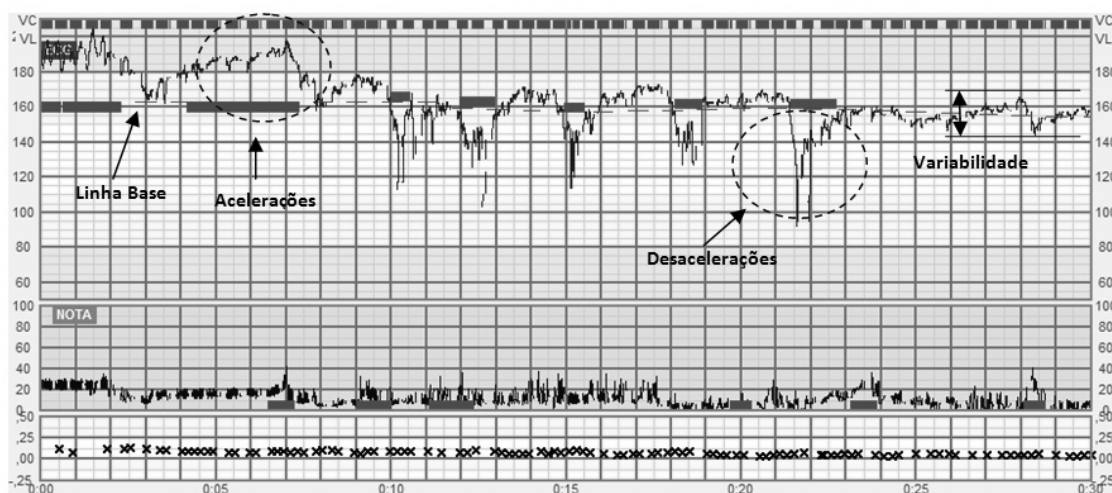


Fig. 2.3: Diferentes características identificadas no traçado CTG.

em 10 minutos, admitindo um traçado médio de 30 minutos) ou com necessidade de uma monitorização mais cuidada, caso apresentem mais de 5 contrações em 10 minutos, sejam estas espontâneas ou induzidas [4].

Podem então diferenciar-se 3 tipos principais de traçados:

Normal Também designado por CTG Tranquilizador, define-se pela linha de base entre 110-160 bpm, a amplitude da variabilidade da FCF entre 5 e 25 bpm, ausência de desacelerações exceto se forem esporádicas ou de muito curta duração, presença de duas ou mais acelerações durante um período de 10 minutos. Uma vez observadas estas acelerações e desde que a variabilidade seja normal e não existam desacelerações, não é necessário continuar o registo.

Não tranquilizador Quando existe falha de algum critério da normalidade. Devem ser investigadas e corrigidas as causas que provocam a anomalia do estado fetal através de uma vigilância contínua e reavaliação a curto prazo. Poder-se-á recorrer a testes mais específicos como a monitorização STAN ou estimulação fetal.

Patológico Associa-se à elevada probabilidade de falta de oxigénio no feto pelo que é necessário a reversão dessa situação ou a rápida intervenção na extração do ambiente intrauterino.

Na figura 2.4 é apresentado um caso em que se procede à análise e classificação do sinal com o sistema Omniview-SisPorto®, descrito mais à frente neste capítulo. Na parte superior podem consultar-se os resultados gráficos do traçado CTG a que se encontra associado um alerta visual indicando o estado fetal. São geradas também informações quantitativas relativamente a características importantes como acelera-

ções, desacelerações, movimentos fetais entre outras que se encontram registadas na parte inferior.

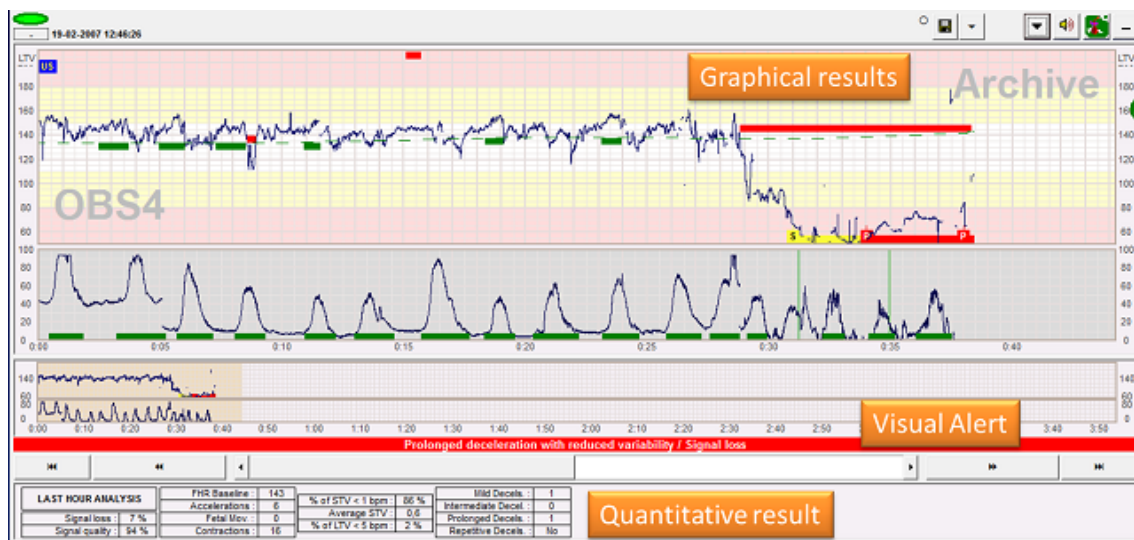


Fig. 2.4: Visualização de sinal CTG analisado e classificado pelo Omniview-SisPorto® com os respetivos resultados quantitativos [30].

As classificações apresentadas são baseadas em estudos reconhecidos e consensualmente aprovados. Podem consultar-se na literatura os parâmetros mais específicos de cada característica [1, 2, 3, 4, 6, 27, 28].

2.1.3 Análise automatizada do CTG

Ao longo das últimas décadas muitos investigadores têm-se focado no desenvolvimento de diferentes algoritmos/métodos que permitam efetuar uma leitura automatizada do traçado de CTG [6, 22, 31].

O facto de a leitura dos traçados ser efetuada visualmente significa estar sujeita a considerável variabilidade intra e inter observadores com as consequentes implicações clínicas. As variações na leitura são diversas podendo ser ao nível da identificação da linha de base da FCF bem como das acelerações ou desacelerações, ou mesmo até na própria interpretação e posterior classificação dos sinais. Desta forma, mesmo que a leitura seja efetuada por clínicos muito experientes e com base nos mesmos critérios de avaliação, o valor clínico de um CTG apresenta um elevado cariz aleatório [1, 8].

A principal e evidente vantagem dos sistemas automáticos é a capacidade de reduzir a subjetividade da análise, uniformizando assim a classificação e interpretação do CTG. Por forma a conseguir-se obter melhores informações sobre o estado do feto, como referido da secção 2.1, a monitorização fetal atualmente combina a avaliação do CTG com o ECG, sendo esta a base da tecnologia *STAN®*, *Neoventa*, *Suécia*. Esta conjugação tem mostrado que é possível reduzir as intervenções desnecessárias

causadas por uma elevada incidência de padrões falso-positivos da FCF e também diminuir os casos de hipóxia do feto.

No entanto, surgem resultados adversos quanto ao uso desta tecnologia, prendendo-se com o facto do erro humano, como sendo a má interpretação do traçado da FCF, atraso na decisão de intervenção por parte do clínico, a não observação das diretrizes clínicas e, em menor número de casos, a não ocorrência ou ocorrência tardia dos eventos ST [8, 4, 23, 26].

Sistema Omniview-SisPorto®

O Omniview-SisPorto® é um sistema central de monitorização fetal e foi desenvolvido, nos últimos 20 anos por investigadores da Faculdade de Medicina da Universidade do Porto (FMUP) em colaboração com o Instituto de Engenharia Biomédica (INEB).

Atualmente comercializado pela Speculum® (Lisboa, Portugal), o programa é usado nos blocos de parto de diversos hospitais em Portugal mas também noutros países como a Dinamarca, a Holanda, o Reino Unido, a Suíça e Israel. Trata-se de um sistema que analisa os sinais fetais habitualmente monitorizados durante o trabalho de parto, detetando alterações associadas à baixa oxigenação fetal e avisando os profissionais de saúde através de alertas sonoros e visuais, emitidos em tempo real [1, 8, 9, 30, 32, 33].

Este sistema permite a visualização de até 25 traçados de CTG e/ou de eventos ST e possui características importantes como a divulgação de alertas visuais e sonoros no decorrer da aquisição do sinal se existir anomalia, o armazenamento de traçados para posterior consulta, impressão ou até mesmo simulações com fins educativos e de formação. A figura 2.5 mostra uma janela de visualização do Omniview-SisPorto® com diferentes simulações. O Omniview-SisPorto® é capaz de identificar situações de irregularidade como, por exemplo, excessivas contrações uterinas, desacelerações prolongadas, variabilidade longa ou curta reduzida e a combinação de eventos ST e características suspeitas do CTG [11, 30].

Alertas Visuais e Sonoros

O Omniview-SisPorto® lidera a tecnologia em análise de CTG. As definições usadas para os alertas são retiradas de publicações reconhecidas e consensuais em interpretação de CTG e eventos ST [2, 8, 11, 30]. Contudo, foram feitas adaptações de acordo com anteriores otimizações e estudos validados, sendo então definidas 6 categorias de alertas que se encontram resumidas na tabela 2.1.

Os alertas “**Attention**” (alerta azul sem som) ocorrem quando as alterações não estão normalmente associadas à hipóxia fetal, como por exemplo a perda de sinal ou o padrão no CTG corresponder à frequência cardíaca materna (FCM).



Fig. 2.5: Visualização Omniview-SisPorto® : Janela do programa com apresentação simultânea de 3 simulações com os respetivos alertas [2].

Os alertas “**Attention Extra**” (alerta visual azul e sonoro) sucedem-se na presença de modificações que podem ou não estar associadas à falta de oxigénio. Podem ocorrer em situações como referidas no alerta anterior, contudo este requer uma maior atenção do profissional para a necessidade de uma possível intervenção.

O alerta “**Reassuring**” (alerta verde), corresponde ao caso em que todos os critérios observados se encontram normais (a linha base da FCF, acelerações, desacelerações e a variabilidade estão presentes). Apenas é processado 20 a 60 minutos após o início da aquisição do traçado.

Os alertas “**Non-Reassuring**” (alerta visual amarelo) definem-se, normalmente, por não se associarem à hipóxia fetal mas poder requerer uma monitorização contínua e talvez mais investigação. Como por exemplo em casos de alterações da linha base da FCF, repetitivas desacelerações de curta duração, ausência de acelerações entre outros.

Os “**Very Non-Reassuring**” (alertas visuais laranjas com som), representam alterações mais preocupantes na FCF que está maioritariamente associado ao risco iminente de hipóxia fetal. Identificam-se pelo facto de existirem características como desacelerações repetitivas e de moderada duração, desacelerações prolongadas e desacelerações curtas com leves eventos ST. É necessário que sejam tomadas medidas para evitar o progresso da situação.

Tab. 2.1: Descrição dos alertas do Omniview-SisPorto® [1, 2, 11].

Alerta	Definição
<i>Attention</i>	Perda de sinal Excessivas Contrações Uterinas Monitorização da Frequência Cardíaca Materna? (Sinal compatível com a FCM) Perda de Sinal ST
<i>Attention Extra</i>	Grande Perda de sinal Monitorização da FCM? (Sinal compatível com a FCM) Desacelerações Monitorização da FCM? Desacelerações muito prolongadas
<i>Reassuring</i>	Todos os parâmetros normais
<i>Non-Reassuring</i>	Bradicardia Basal Mudança da Linha de Base Taquicardia Fetal Ausência de Acelerações Desacelerações Desacelerações Repetitivas
<i>Very Non-Reassuring</i>	Várias Desacelerações Repetitivas Desacelerações prolongadas Desacelerações Repetidas e eventos ST
<i>Pre-Terminal</i>	Reduzida Variabilidade Longa Desacelerações Repetitivas e Reduzida Variabilidade Longa Taquicardia, Desacelerações e eventos ST Repetitivas Desacelerações e eventos ST Desacelerações e eventos ST Várias Desacelerações Repetidas e eventos ST Desacelerações prolongadas e eventos ST Reduzida Variabilidade Curta Desacelerações Repetitivas e Reduzida Variabilidade Curta Desacelerações Repetitivas muito prolongadas Prolongadas Desacelerações e Reduzida Variabilidade

Os alertas “**Pre-Terminal**” (alertas vermelhos com som) são aqueles que traduzem as situações de maior gravidade. Fortemente ligados á hipóxia fetal moderada ou severa, identificam-se pela reduzida variabilidade de curta e longa duração, repetitivas ou prolongadas desacelerações com reduzida variabilidade dos eventos ST, taquicardia entre outros. Requer uma rápida e eficaz intervenção para que se efetue a reversibilidade rápida das causas de estado fetal crítico, sendo uma das soluções o parto imediato [1, 2, 11, 29].

De seguida são apresentados alguns exemplos de alertas que demonstram a correta avaliação do estado fetal. Na figura 2.6 é apresentado um caso em que o excesso de contrações uterinas deve motivar uma cuidadosa monitorização pois pode sugerir uma baixa concentração de oxigénio no sangue arterial que se relaciona com alterações na FCF.

No caso de um alerta muito suspeito, representado pela figura 2.7, pode verificar-se que corresponde a desacelerações muito repetitivas podendo estas indicar algum risco grave para o feto e para tal deve evitar-se grande espera para a intervenção adequada.

O alerta patológico, figura 2.8, tem indicações que se associam a uma elevada probabilidade de hipóxia fetal e deve provocar uma imediata ação de profissionais

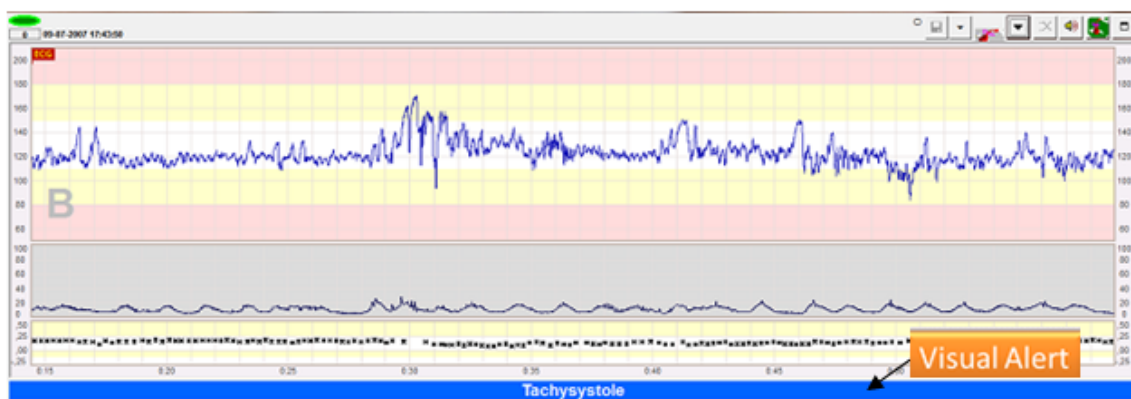


Fig. 2.6: Traçado CTG representativo do alerta *Attention* [30].

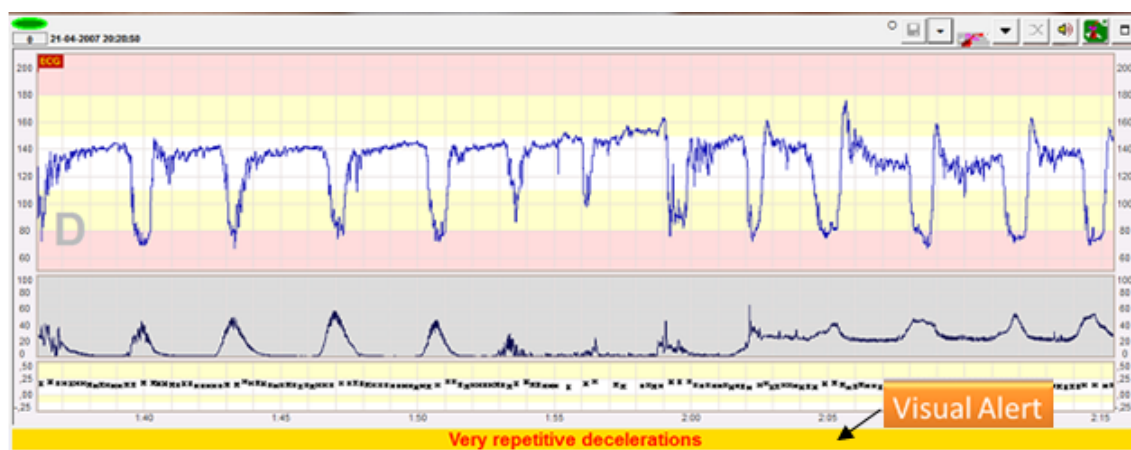


Fig. 2.7: Sinal CTG representativo do alerta *Very Non-Reassuring* com repetitivas desacelerações [30].

para a reversão da causa do sofrimento ou então a remoção rápida do feto do ambiente intrauterino.

Em todos estes casos de análise do CTG é necessário ter-se em conta, entre outros aspetos que possam influenciar o parto, a idade gestacional, possíveis drogas administradas ou os anteriores padrões de CTG.

2.2 Redes Neurais Artificiais

As RNAs têm sido amplamente estudadas e vindo a revelar-se importantes para os investigadores de IA. Também conhecidas como modelos conexionistas, são por definição sistemas computacionais estruturados numa aproximação à computação baseada em ligações.

Na tentativa de imitar o sistema nervoso, este processo utiliza um variado número de estruturas idênticas e interconectadas que se designam por neurónios. Possuem a predisposição de guardar conhecimento experimental e torná-lo disponível para uso. Esse conhecimento é adquirido pelas RNAs através do meio envolvente, por

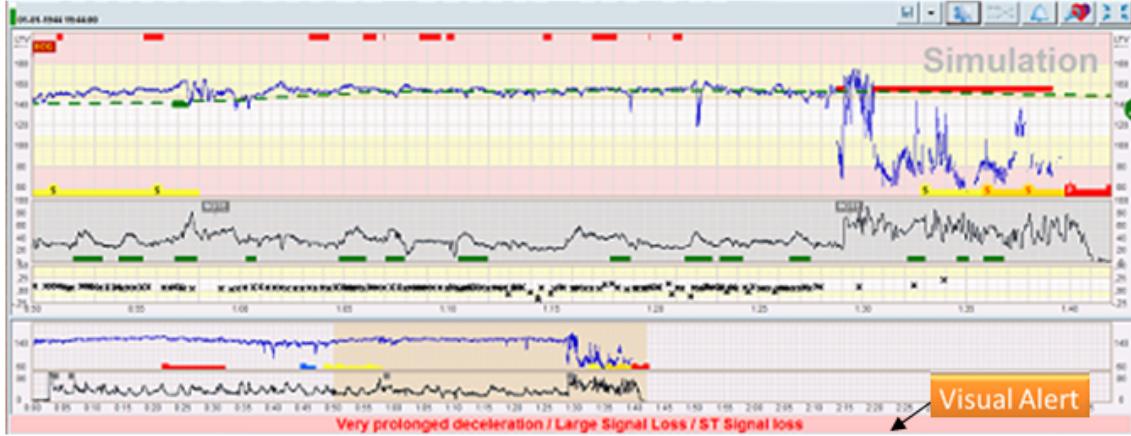


Fig. 2.8: Exemplo de um traçado patológico com o respetivo alerta do tipo *Pre-Terminal* [30].

processos de aprendizagem, e armazenado nas conexões interneuronais designadas por pesos sinápticos.

Uma das principais características das RNAs é a capacidade de se conseguirem adaptar ao meio envolvente, isto é, a diferentes casos de estudo, influenciando assim positivamente a estabilidade e robustez do sistema. Por conseguinte, esta técnica de IA tem sido cada vez mais utilizada como método válido para solucionar problemas complexos e de difícil resolução, assumindo-se como uma valiosa ferramenta para diversas aplicações [13, 34, 35, 36].

2.2.1 Modelo do Neurónio Artificial

O elemento fundamental de uma rede neuronal é o neurónio que, como referido em 2.2, inspira-se na biologia dos neurónios do sistema nervoso humano. Um diagrama exemplificativo de um destes tipos de neurónios está representado na figura 2.9. Podemos verificar que é composto por três componentes principais:

1. Um conjunto de sinais de entrada para o somatório, cada um deles com um peso sináptico associado. A conexão entre o sinal de entrada x_j e o neurónio k tem o peso de w_{kj} , podendo estes tomar qualquer valor positivo ou negativo.
2. Um operador que efetua a soma do valor do viés b_k com as entradas ponderadas pelos seus respetivos pesos sinápticos w_{kj} .
3. Uma função de ativação, responsável por limitar a amplitude da saída y_k do neurónio. As amplitudes mais comuns para o neurónio de saída são os intervalos fechados de $[0,1]$ ou $[-1,1]$.

O conjunto de todas as conexões dos neurónios de uma RNA forma o vetor de pesos, cuja dimensão é definida pela arquitetura de rede. Durante o processo de aprendizagem esse vetor é ajustado de forma a atingir o objetivo desejado [34, 35].

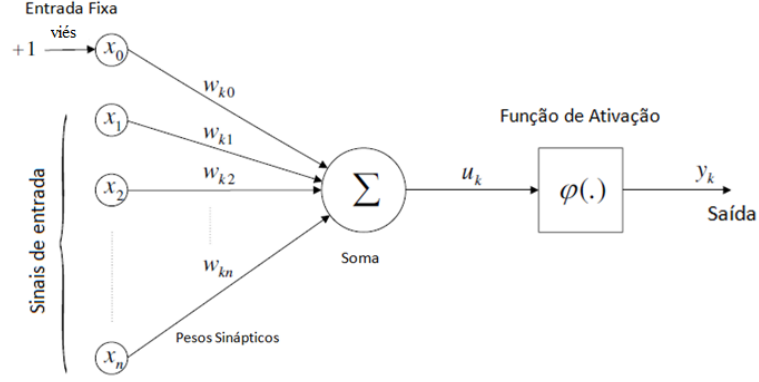


Fig. 2.9: Modelo de um neurónio artificial constituído por um conjunto de pesos, uma unidade de soma e uma função de ativação.

Em termos matemáticos, o neurónio k pode ser definido por:

$$v_k = \sum_{j=1}^n w_{kj} x_j, \quad (2.1)$$

onde x_1, x_2, \dots, x_n são os sinais de entrada, $w_{k1}, w_{k2}, \dots, w_{kn}$ representam os pesos dos neurónios e v_k é a saída do combinador linear.

Como se pode ver na figura 2.9, existe um termo designado viés, b_k , que se trata de uma ligação extra que estabelece uma tendência ou inclinação no processo computacional, isto é, adiciona uma constante para que se estabeleçam as corretas condições operacionais. A equação 2.2 traduz a soma de v_k com b_k . O resultado desta soma, u_k , será o argumento da função de ativação, $\varphi(\cdot)$,

$$u_k = v_k + b_k. \quad (2.2)$$

O valor de saída, y_k , será dado pela equação:

$$y_k = \varphi(u_k). \quad (2.3)$$

Finalmente, efetuando as respetivas substituições, a equação geral que traduz a computação de um neurónio deste tipo é:

$$y_k = \varphi\left(\sum_{j=1}^n w_{kj} x_j + b_k\right). \quad (2.4)$$

2.2.2 Funções de Ativação

A função de ativação, representada por $\varphi(u)$, restringe a amplitude da saída de um neurónio num intervalo finito. Na figura 2.10 podemos ver as três principais funções de ativação.

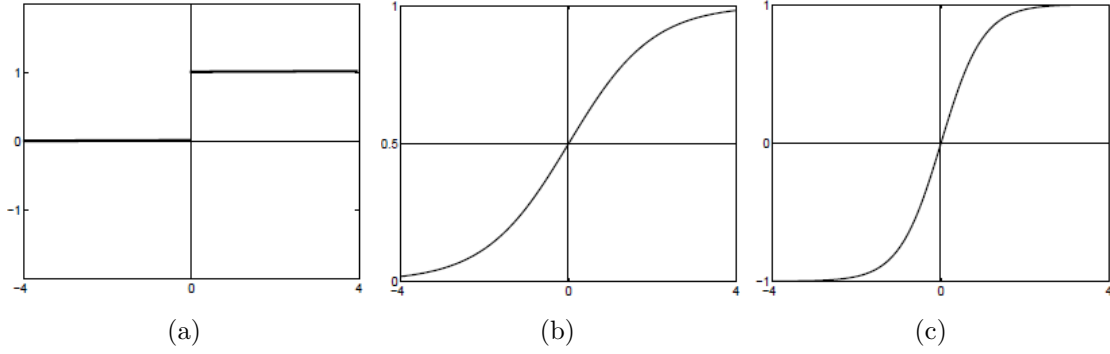


Fig. 2.10: Funções de ativação mais utilizadas: (a) Função Degrau; (b) Função Logística Sigmoidal com $a=1$; (c) Função Tangente Hiperbólica com $a=1$ [36].

A função Degrau, também conhecida como Função de *Heaviside* ou de *threshold*, é definida por:

$$\varphi(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0. \end{cases} \quad (2.5)$$

Com esta função de ativação a saída da RNA é 1 se u_k não for negativa e 0 quando for. O modelo do neurónio com a função degrau foi primeiramente introduzido por *McCulloch-Pitts* e posteriormente desenvolvido por *Rosenblatt*, que ficou então conhecido como perceptrão.

A função sigmoide com a sua forma típica de um S é aquela que é mais utilizada em aplicações de RNAs. Esta função monótona crescente descreve um balanceamento entre o comportamento linear e não linear [34]. Duas destas funções sigmoide são a função logística, 2.6, e a função tangente hiperbólica, 2.7:

$$\varphi(u) = \text{sig}(u) = \frac{1}{1 + e^{-au}}, \quad (2.6)$$

$$\varphi(u) = \tanh(u) = \frac{e^{au} - e^{-au}}{e^{au} + e^{-au}}. \quad (2.7)$$

O parâmetro a de ambas as funções controla o declive da tangente à curva na origem. As saídas das RNAs usando a função logística sigmoide e \tanh encontram-se no intervalo $[0,1]$ e $[-1,1]$, respetivamente [34].

Alguns algoritmos de treino, como é o caso do algoritmo de *backpropagation*, exigem que a função de ativação seja diferenciável e monótona. A função sigmoide contempla estas características sendo diferenciável em todo o seu domínio, tendo como derivadas (para $a=1$):

$$\text{sig}'(u) = \frac{e^{-u}}{(1 + e^{-u})^2} = \text{sig}(u)(1 - \text{sig}(u)), \quad (2.8)$$

e

$$\tanh'(u) = \frac{4}{(e^u + e^{-u})^2} = 1 - \tanh^2(u). \quad (2.9)$$

Funções como a $\tanh(u)$, que produzem quer valores positivos quer negativos, tendem a realizar um treino mais rápido da rede do que funções que produzam apenas valores positivos, como é o caso da logística. Contudo a função logística permite que, posteriormente, a saída possa ser interpretada como probabilidades fornecendo mais do que uma simples classificação [35, 36].

2.2.3 Arquiteturas de rede

Uma RNA é formada por neurónios conetados entre si através de pesos e é na fase de treino que estes parâmetros são modificados de forma a obter o desempenho correto da rede. A arquitetura restringe o tipo de problema no qual a rede possa ser utilizada e dependem portanto do número de camadas, número de neurónios em cada camada, do tipo de conexão entre os nós e da conectividade existente na rede. Existem diferentes tipos de RNAs contudo a grande distinção é feita entre:

- Redes *FeedForward* com apenas uma camada, multi-camadas e Redes de *Kohonen*.
- Redes *FeedForward* e *Feedback*, Redes de *Hopfield* e Redes Recorrentes.

Redes FeedForward

As redes FF são organizadas em camadas devido à inexistência de ciclos, visto que as conexões existentes são sempre unidirecionais, ou seja, a informação é processada apenas da entrada para a saída (e nunca o contrário). A arquitetura mais simples de rede é a rede perceptrão de apenas uma camada, figura 2.11, que é constituída apenas pelos nós de entrada e a camada de neurónios de saída. De referir que a camada de entrada não é contabilizada como sendo uma camada da RNA por não se efetuar qualquer tipo de computação [34, 37].

Podem-se obter outras arquiteturas mais complexas através da adição de camadas na construção da rede. Estas redes são designadas por perceptrão com múltiplas camadas (MLP) em que a rede pode ter uma ou mais camadas intermédias tendo como função intervir entre a camada de entrada e de saída da rede. A figura 2.12 representa um exemplo de uma MLP com apenas uma camada escondida. Adicionando-se uma ou mais camadas intermédias aumenta-se a capacidade da rede em modelar problemas mais complexos. No entanto, o tempo de aprendizagem da rede aumenta e requer um maior poder computacional de processamento e armazenamento da rede [34].

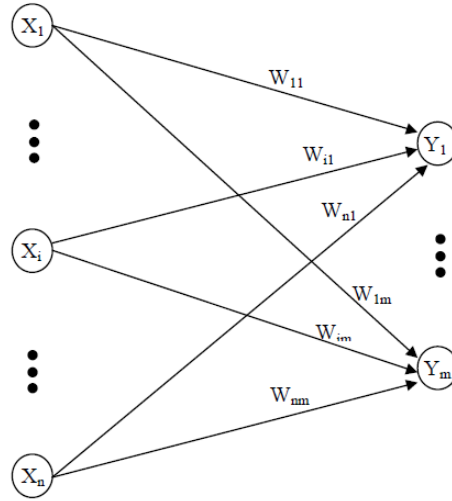


Fig. 2.11: Rede neuronal do tipo *feedforward* de apenas uma camada.

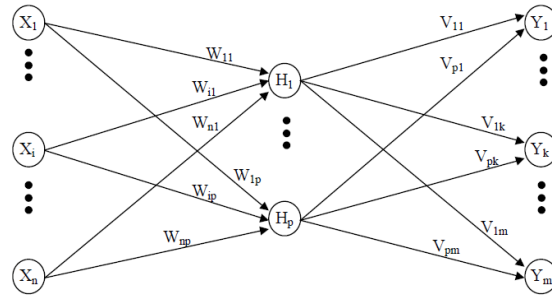


Fig. 2.12: Representação de uma MLP com uma camada escondida.

As saídas dos neurónios de cada camada da rede são utilizadas como entradas para a camada seguinte desta e uma das formas de aprendizagem desta rede é através do algoritmo de *backpropagation*, que será explicada mais à frente neste trabalho.

Neste trabalho será esta a arquitetura de rede adotada para a classificação e será utilizada a notação de $[n_h: n_out]$ sendo n_h o número de neurónios na camada escondida e n_out o número de neurónios na camada de saída.

Existe também um grande interesse no uso de MLP para previsão de séries temporais que se deve à característica de aproximação universal desta arquitetura. A capacidade de modelação de um processo a partir de uma série temporal faz com que a rede adquira um comportamento dinâmico designando-se por rede neuronal dinâmica. A MLP é tornada sensível à estrutura temporal das séries de informação através da introdução de memória na rede e uma forma simples de o conseguir é através de atrasos na entrada (*time delays*) ou a recorrência das redes. A FTDNN é um exemplo de uma MLP dinâmica não-recorrente e pode ver-se na figura 2.13 o esquema da sua arquitetura [16]. Este foi o tipo de rede selecionado para se efetuar a previsão da série CTG, apresentada mais à frente nesta dissertação.

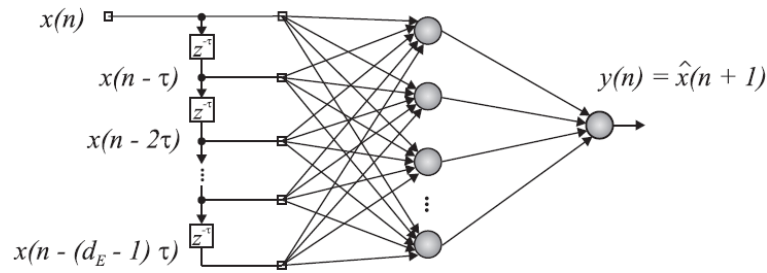


Fig. 2.13: Arquitetura genérica de uma rede FTDNN de uma camada escondida [16].

Redes Recorrentes

As redes recorrentes são aquelas que contêm conexões sinápticas realimentadas, ou seja, prevê ligações entre elementos da mesma camada e/ou com elementos de camadas anteriores, ou ainda para a própria camada. A recorrência é um tipo de memória de curta duração que permite à rede adquirir um comportamento dinâmico com propriedades de memória [34, 16]. Um exemplo deste tipo de redes encontra-se representado na figura 2.14.

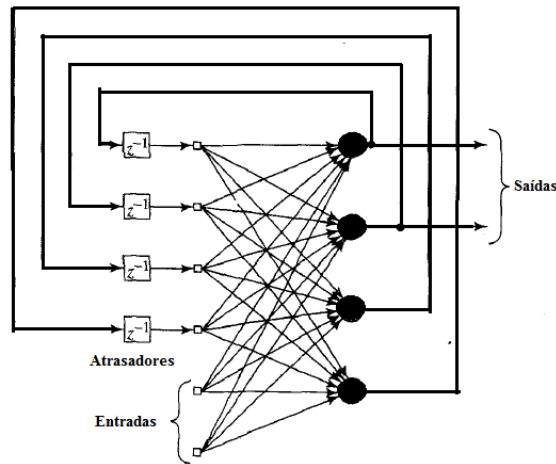


Fig. 2.14: Esquema representativo de uma RNA recorrente.

A principal diferença entre as redes recorrentes e MLPs dinâmicas não-recorrente, como a FTDNN, está no processamento que a realimentação realiza sobre a informação passada, tornando esta arquitetura de rede mais indicada para problemas como a previsão [16, 38, 39, 40].

2.2.4 Processos de Aprendizagem

Considera-se que uma das propriedades mais importantes das RNAs é a sua capacidade de aprender com base no meio em que está inserida. Segundo Haykin [34], o objetivo da aprendizagem consiste no ajuste dos valores dos pesos sinápticos da rede, de acordo com algumas regras, de forma a alcançar o conjunto de saídas desejadas

ou aproximadas com um intervalo de erro estabelecido.

O conjunto de regras sobre o qual a rede baseia a sua aprendizagem chama-se de algoritmo de aprendizagem e as cinco principais regras básicas de aprendizagem são: com correção de erros (*Error Correction Learning*), baseada na memória, Aprendizagem *Hebbian*, Competitiva e de *Boltzmann* [34].

A aprendizagem pode também ser efetuada tendo em conta três paradigmas básicos de adaptação de parâmetros do sistema: aprendizagem por reforço, métodos supervisionados e não supervisionados. No primeiro caso a resposta correta não é apresentada à rede, isto é, apenas se fornece uma indicação sobre se a resposta da rede está ou não correta ou errada, tendo a rede de usar esta informação para melhorar o seu desempenho. Na aprendizagem supervisionada cada padrão de treino colocado na entrada da rede está sempre associado ao seu par correspondente na saída. Durante o treino, existe sempre a informação de cada padrão de entrada e o respetivo de saída que representa a resposta pretendida e que a rede deverá aprender.

Por sua vez, na aprendizagem não supervisionada não existe qualquer tipo de informação do erro entre a saída obtida e a saída pretendida, uma vez que não é conhecida a saída desejada. Durante a aprendizagem tentam-se descobrir similaridades e dissimilaridades nos padrões de treino, de maneira a agrupá-los em classes de características mais ou menos comuns. Na figura 2.15 apresenta-se, de forma esquemática, os modelos de aprendizagem supervisionada e não supervisionada.

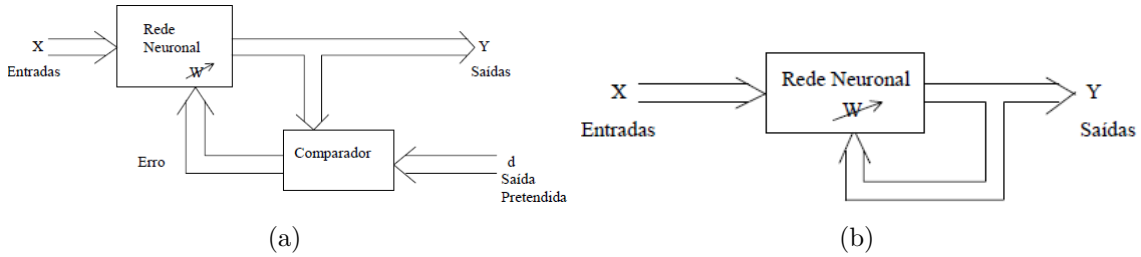


Fig. 2.15: Modelos representativos de aprendizagens: (a) Supervisionada; (b) Não supervisionada [35].

Error-correction learning

Para ilustrar este processo de aprendizagem, consideremos o esquema representado na figura 2.16, que pode conter uma ou mais camadas, apenas um neurónio de saída e a resposta desejada designa-se por $d_k(n)$. O termo $y_k(n)$ representa o sinal de saída do neurónio k à iteração n .

O erro entre o objetivo desejado $d_k(n)$ e o sinal de saída obtido $y_k(n)$ pode ser definido como:

$$e_k(n) = d_k(n) - y_k(n). \quad (2.10)$$

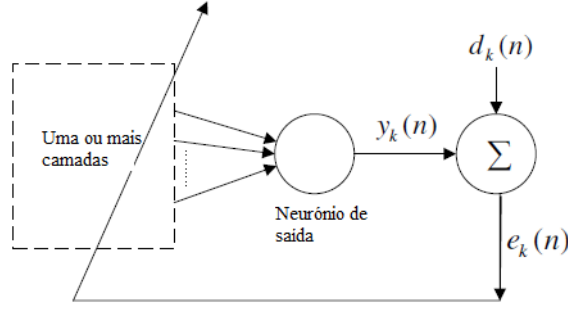


Fig. 2.16: Esquema Representativo da *Error-correction learning*.

O processo de aprendizagem efetua o ajustamento dos pesos da RNA de forma a que a saída do neurónio se torne cada vez mais próxima da saída desejada. Este ajuste pode ser feito, por exemplo, através da função de minimização do erro quadrático $E(n)$ definida em termos do sinal erro $e_k(n)$ como:

$$E(n) = -\frac{1}{2}e_k^2(n). \quad (2.11)$$

A aprendizagem é executada através do ajuste dos pesos da RNA até se obter o mínimo valor de $E(n)$. Esta minimização da função erro é alcançada através da regra Delta ou regra de *Widrow-Hoff*. Consideremos uma RNA simples com uma função de ativação linear, sendo $w_{kj}(n)$ o valor do peso w_{kj} da conexão entre o neurónio k e o elemento $x_j(n)$ do vetor de entrada $x(n)$ no instante n . De acordo com a regra Delta, o ajuste $\Delta w_{kj}(n)$ aplicado ao peso sináptico $w_{kj}(n)$ no instante de tempo n é definido por:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n), \quad (2.12)$$

onde η é uma constante positiva que determina a taxa de aprendizagem no processamento de um intervalo de aprendizagem para outro. O parâmetro η é conhecido como taxa de aprendizagem.

A atualização dos pesos, w_{kj} , após computação de $\Delta w_{kj}(n)$ é obtido a partir de:

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n). \quad (2.13)$$

A taxa de aprendizagem é um dos parâmetros mais importantes das RNA e deve ser escolhido cuidadosamente para que a convergência do processo iterativo de aprendizagem seja alcançada.

Na aprendizagem de correção de erros o ajuste dos pesos pode ser feito através da propagação dos sinais de erro para trás, camada a camada [34, 36]. Isto é feito com base no algoritmo de treino de retro-propagação, que é apresentado seguidamente.

Algoritmo retro-propagação

O algoritmo de retro-propagação é um exemplo de aprendizagem supervisionada que pode ser visto como uma generalização da regra Delta para redes neurais de múltiplas camadas.

Vários investigadores interessaram-se por este algoritmo como *Williams* e *McClelland* em 1986, *Minsky* e *Papert* em 1988, entre muitos outros, tentando alcançar diferentes objetivos. Este algoritmo é baseado na correção de erros e o seu objetivo principal é encontrar na superfície de erro um mínimo global, ou seja, valores para os pesos sinápticos que minimizam o erro total da rede [34].

Visto que utilizamos este método é apresentada de seguida uma breve explicação. Na figura 2.17 encontra-se representado um fluxo de sinais simplificado de uma RNA de apenas uma camada com função de ativação sigmoide.

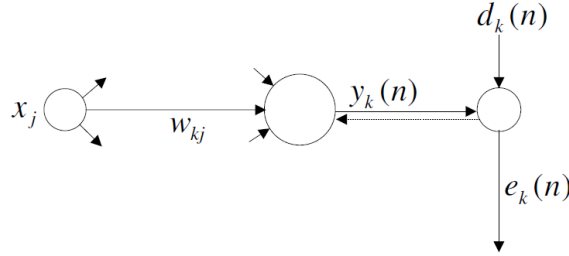


Fig. 2.17: Fluxo do sinal e retro-propagação do erro (tracejado) numa rede perceptrão simples.

Admitindo que $e_k(n) = d_k(n) - y_k(n)$ é o erro entre a saída do neurónio k e o alvo pretendido n , para atualizar os pesos da RNA utiliza-se então a regra Delta, mais precisamente, o algoritmo de descida do gradiente. Admitindo também a função definida em 2.11, a regra Delta para o peso sináptico w_{kj} será:

$$\Delta w_{kj}(n) = -\eta \frac{\partial E}{\partial w_{kj}}. \quad (2.14)$$

De notar que E é uma função de $y_k(n)$ que, por sua vez, é uma função de u_k e consequentemente depende dos pesos sinápticos (Equação 2.3). A fim de calcular 2.14, aplicando a regra de derivação obtém-se:

$$\Delta w_{kj}(n) = -\eta \frac{\partial E}{\partial w_{kj}} = \eta \delta_k(n) x_j(n), \quad (2.15)$$

onde o gradiente $\delta_k(n)$ é definido por [34, 36]:

$$\delta_k(n) = e_k(n) \varphi'(u_k(n)). \quad (2.16)$$

Resumidamente, o algoritmo de aprendizagem tem por base a utilização de pares de entrada e saída para ajustar os pesos da rede. A frequência com que estes ajustes

são feitos depende do tipo de treino adotado pela rede, podendo ser:

- Modo incremental onde a atualização é realizada após cada exemplo apresentado;
- Por ciclo (*batch* ou lote) onde a atualização é realizada após todos os exemplos do conjunto de treino serem apresentados à rede

Podem definir-se duas fases principais: a *Forward* e a *Backward*. Na fase *Forward*, a entrada é apresentada à primeira camada da RNA. Após os nós da camada consecutiva calcularem os sinais de saída, estes irão servir como entrada para a definição das saídas produzidas pelos nós da camada seguinte e assim sucessivamente. Por fim, a saída obtida é então comparada com aquela desejada.

Por sua vez, na fase *Backward* os nós da camada atual ajustam os pesos de forma a reduzir o erro. O cálculo do erro de um nó das camadas escondidas não é um processo muito trivial, sendo que este é o cálculo feito utilizando os erros dos nós da camada seguinte ligados à camada escondida em questão, ponderada pelos pesos das conexões entre elas. Esta fase dá-se a partir da última camada, até chegar à camada de entrada. Note-se que a retro-propagação dos erros através da rede permite o correto ajuste dos pesos sinápticos entre todas as camadas do modelo.

Embora ainda não haja critérios com sucesso para terminar o processo de aprendizagem, existem os designados critérios de paragem como são exemplo o erro mínimo ou o pretendido atingido, quando a variação do erro for infinitamente pequena sendo insuficiente para melhorar a *performance* da rede ou atingir-se o número máximo de iterações durante o treino. Neste último caso, a RNA não aprende mais, isto é, memoriza os padrões de treino e portanto o erro diminui no conjunto de treino, embora para os casos de teste piore o desempenho, perdendo capacidade de generalização. Este fenómeno designa-se por *overfitting* e pode ser evitado escolhendo outras arquiteturas de rede, definindo outros critérios de paragem que façam com que a rede seja menos complexa e impeça de memorizar padrões [13, 17, 34, 35].

Taxa de aprendizagem

Já foi referido que a taxa de aprendizagem é um dos fatores mais importantes no treino da RNA. Admitindo o algoritmo de retro-propagação com descida do gradiente sabe-se que em cada iteração é encontrado o erro cometido. Ao admitir-se um valor para este parâmetro baixo, as alterações dos valores dos pesos sinápticos será também baixa e consequentemente não se refletirá muito na superfície do erro. O que acontece é que a aprendizagem da RNA torna-se muito lenta. No entanto, ao selecionar-se uma taxa de aprendizagem muito alta fará com que o ajuste dos pesos sinápticos seja muito brusca provocando uma trajetória muito “caótica” na

superfície do erro. Desta forma é extremamente importante a seleção de um valor que evite as situações mencionadas anteriormente [36].

2.2.5 Medida de *performance* de classificadores

A análise da *performance* dos classificadores é fundamental para se perceber o comportamento do algoritmo e os consequentes resultados da classificação. A partir de um conjunto de dados e de um algoritmo a dúvida que surge é de que forma se irá estimar a taxa de erro do classificador.

Existem para tal diferentes funções de erro e índices de *performance* que nos auxiliam na análise das classificações e estes serão apresentados resumidamente de seguida.

Funções de erro

Tal como referimos em secções anteriores (2.2.4), o ajuste dos pesos de uma RNA é importante para minimizar a função de erro⁴, definida pelo sinal erro $e=d-y$.

Existem inúmeras funções de erro usadas em RNA sendo a função do erro quadrático médio (*Mean Squared Error*, MSE) a mais utilizada em problemas de classificação com MLPs.

A MSE, como outros índices baseados no erro quadrático, é também usada para avaliar a *performance* da rede. Estes índices são portanto usados para comparar a *performance* de diferentes soluções da RNA. Na tabela 2.2 é apresentado o conjunto de índices de *performance* mais utilizado [36]:

O desvio padrão do erro,

$$\sigma_e = \frac{1}{N-1} \sqrt{\sum_{i=1}^N (e_i - m_e)^2}, \quad (2.17)$$

também pode ser utilizado como medida de *performance* em problemas de regressão.

A análise ROC (*Receiver Operating Characteristic*) teve origem na teoria de decisão estatística com o objetivo de avaliar a deteção de sinais em radares e atualmente tem sido muito aplicada em medicina, radiologia e psicologia numa grande variedade de testes de diagnóstico [13, 41]. A curva ROC é comumente utilizada para apresentar resultados de problemas binários e pretende mostrar de que forma o número de exemplos com classificações positivas corretas varia em relação ao número de classificações negativas incorretas. A decisão tomada pelo classificador pode ser representada através de uma estrutura conhecida como matriz de confusão (ou ta-

⁴ Também conhecidas como *cost functions*, são usadas na optimização e relacionadas com o problema de forma a encontrar a solução ótima para um problema em particular [36].

Tab. 2.2: Exemplos de índices de *performance*.

Mean Squared error	$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2$
Error Mean	$me = \frac{1}{N} \sum_{i=1}^N e_{-i}$
Absolute Error Mean	$m e = \frac{1}{N} \sum_{i=1}^N e_{-i} $
Relative Error	$E_{rel} = \frac{1}{N} \sum_{i=1}^N \frac{e_i}{x_i}$
Root Mean Square Error	$E_{RMS} = \sqrt{MSE}$

bela de contingência), exemplificada na figura 2.18. Carateriza-se por ser de fácil interpretação e podem considerar-se quatro categorias [41, 42]:

- **Verdadeiros Positivos (VP)** - exemplos positivos corretamente etiquetados como positivos (acerto);
- **Falsos Negativos (FN)** - exemplos positivos incorretamente etiquetados como negativos (erro);
- **Verdadeiros Negativos (VN)** - exemplos negativos corretamente etiquetados como negativos (acerto);
- **Falsos Positivos (FP)** - exemplos negativos incorretamente etiquetados como positivos (erro).

De acordo com a figura 2.19, pode assumir-se que a curva ROC é obtida através de valores de sensibilidade, no eixo das ordenadas, e o complemento da especificidade, no eixo das abcissas (que serão definidas à frente), ou seja, Taxa de Falsos Positivos e Taxa de Verdadeiros Positivos, respetivamente. Para que se considere um classificador perfeito o resultado desejado deve ser correspondente à linha horizontal no topo do gráfico, o que dificilmente é uma situação alcançável.

Para se considerar uma boa classificação, a curva ROC obtida deve posicionar-se entre a linha diagonal e a perfeita, onde quanto maior a distância em relação à linha diagonal melhor será o sistema. Pode ainda dar-se o caso de um sistema apresentar a respetiva curva ROC abaixo da diagonal. Neste caso, para se obter algum sucesso

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP	FP
	negativos	FN	VN

Fig. 2.18: Representação de uma matriz de confusão.

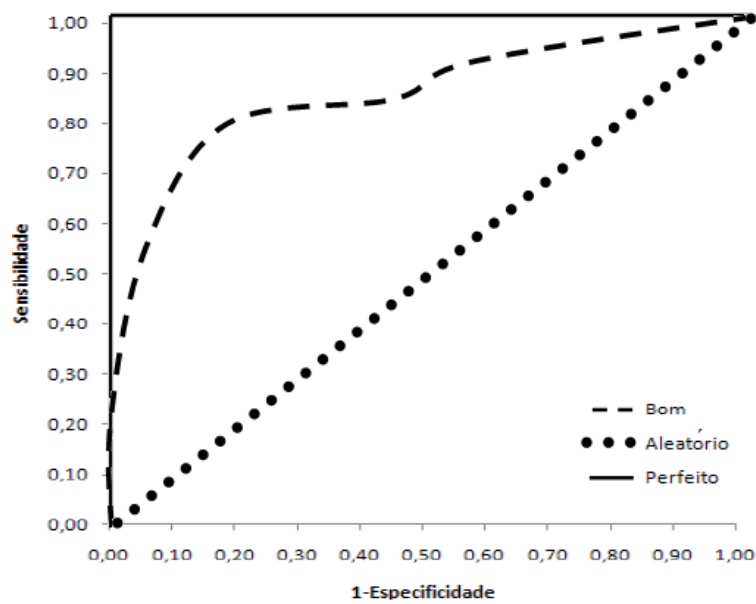


Fig. 2.19: Curva ROC.

do sistema pode inverter-se as suas saídas e consequentemente a curva também o será [13].

De forma a se conseguir efetuar uma comparação entre sistemas é utilizada como medida a área sob a curva (AROC), que pode ser obtida por métodos de integração numérica e, teoricamente, quanto maior for a AROC melhor será então o sistema.

Precisão e *Recall*

A medida precisão diz respeito à proporção de classificações corretas, sendo que esta medida classifica-se como suscetível a desbalanceamentos do conjunto de dados e pode facilmente induzir a uma conclusão errada sobre o desempenho do sistema [13].

Como foi referido anteriormente, estas medidas de *performance* podem ser definidas a partir da matriz de confusão. Desta forma, temos que:

$$Precisão = \frac{VP}{VP + FP}, \quad (2.18)$$

e

$$Recall = \frac{VP}{VP + FN}. \quad (2.19)$$

As curvas *Precision-Recall* têm sido utilizadas em casos de classificação com grande desigualdade no conjunto de dados, ou seja, quando os dados não estão distribuídos homogeneamente pelas classes, evidenciando-se classes com a maior parte dos dados e outras com poucos. Admite-se que estas curvas dão mais informação acerca da *performance* do algoritmo em uso e revelam-se portanto alternativas às curvas ROC.

Na figura 2.20 está representado um exemplo do mesmo modelo de aprendizagem, de um conjunto de dados relativo à deteção de cancro que pretende destacar a diferença visual entre os espaços ROC e PR [42].

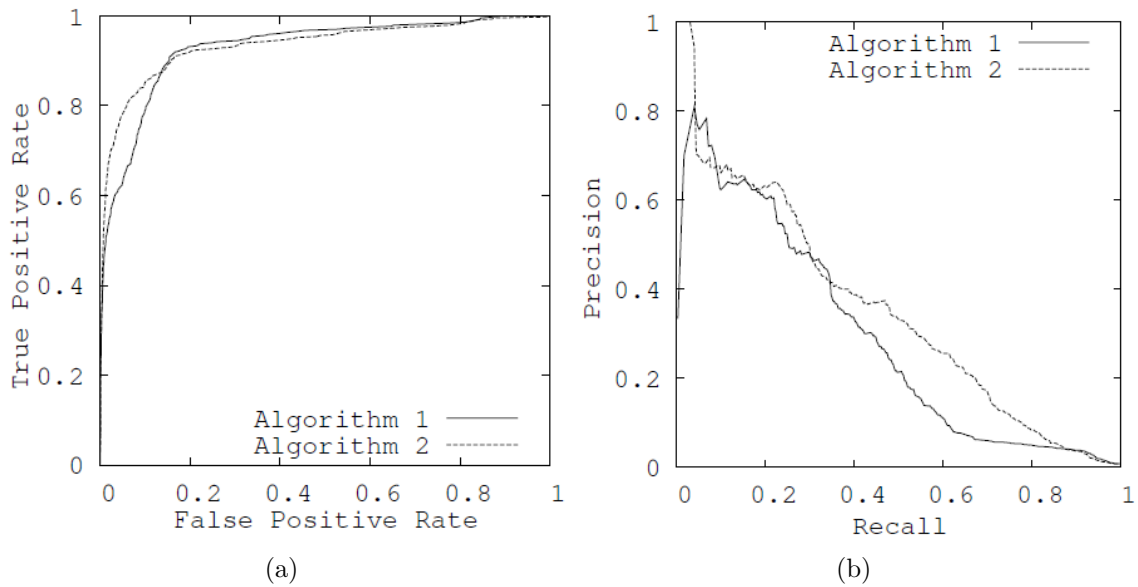


Fig. 2.20: Diferença entre a Comparação de algoritmos usando a curva ROC e PR: (a) Comparação no espaço ROC; (b) Comparação no espaço PR [42].

Na interpretação da ROC o objetivo é que os resultados se situem no canto superior esquerdo mas com o gráfico PR a localização é diferente tendo de se situar no canto superior direito. De acordo com o exemplo, a curva PR demonstra que os resultados não são assim tão satisfatórios, mostrando que ainda há um vasto espaço para melhoria [42].

Sensibilidade e Especificidade

A Sensibilidade⁵ trata-se de uma medida estatística que diz como um teste de classificação binária identifica corretamente uma determinada condição. Na área médica, num ensaio clínico de determinação de uma patologia, esta medida relaciona-se com a probabilidade da pessoa que tem a doença ter o teste positivo. Ou seja, é a proporção de verdadeiros positivos de todos os casos positivos na população [36].

$$Sensibilidade = \frac{VP}{VP + FN}. \quad (2.20)$$

Por sua vez, a medida estatística Especificidade⁶ diz como um teste de classificação binária classifica corretamente casos que não pertencem a essa classe. Seguindo a lógica do exemplo anterior, esta medida é a probabilidade de, se a pessoa não tem a doença, o teste dar o resultado negativo. Trata-se então da proporção de verdadeiros negativos de todos os casos negativos na população [13, 36]:

$$Especificidade = \frac{VN}{VN + FP}. \quad (2.21)$$

Exatidão

A exatidão (ou *accuracy*) serve também como medida de *performance* dos classificadores e foi esta a que se teve em consideração para este trabalho. Está intimamente relacionada com a taxa de erro da classificação e os resultados da classificação com esta medida estatística relacionam-se da forma como mostra a equação 2.22, cuja quantidade de classificações corretas é dividida pelo valor total de exemplos no conjunto. De notar que a exatidão da classificação pode também ser calculada a partir da matriz de confusão:

$$Exatidão = \frac{VP + VN}{VP + VN + FP + FN}. \quad (2.22)$$

Esta medida pode igualmente ser escrita em termos do erro de previsão, ou seja, a diferença entre o valor efetivo e o valor previsto. A Taxa de Erro, TE, ($Exatidão = 1 - Taxa de Erro$), pode ser definida como:

$$TE = \frac{FP + FN}{VP + VN + FP + FN}. \quad (2.23)$$

Num exemplo clínico, a exatidão do diagnóstico é afetada pela prevalência da doença. Com a mesma sensibilidade e especificidade, a exatidão do diagnóstico de

⁵ Também conhecida como *Recall* ou Taxa de Verdadeiros Positivos

⁶ Nas curvas ROC é utilizado graficamente, no eixo das abcissas, o complemento da especificidade que se designa por *False Positive Rate* (*FPR*) ou Taxa de Falsos Positivos caracterizada por $FPR = \frac{FP}{VN + FP}$.

um teste em particular aumenta à medida que a prevalência da doença diminui o que requer um tratamento de dados cuidadoso.

Contudo, não significa que se se considerar uma população com baixa prevalência da doença o teste obtenha um melhor resultado. Apenas revela que se obtém classificações mais corretas, o que deve ser conjugado com outras medidas de *performance* para se poder ter uma avaliação completa [43].

Balanced Error Rate

A Taxa de Erro Equilibrada ou *Balanced Error Rate*, BER, representa-se como a média dos erros em cada classe, sendo dada por:

$$BER = \frac{1}{2} \left(\frac{VP}{VP + FN} + \frac{VN}{VN + FP} \right). \quad (2.24)$$

Classificação de CTG usando FFNN

O presente capítulo inicia-se com uma breve descrição de alguns trabalhos realizados sobre classificação com RNA mais concretamente com a aplicação desta técnica de IA na classificação de CTGs. Seguidamente, são apresentados os conjuntos de dados utilizados neste estudo e feita a descrição e explicação do algoritmo desenvolvido. Por fim, são apresentados e explicados os resultados obtidos da classificação através do algoritmo proposto.

3.1 Estado da Arte

A análise do sinal CTG apresenta dois problemas principais: um está relacionado com os métodos de classificação usados para gerar o alarme de risco durante a gravidez. Muitas vezes estes não são adequados por apenas descreverem aspetos morfológicos do sinal, não tendo relações entre o conjunto dos parâmetros e o potencial estado fetal patológico. Outro aspeto é o facto de algumas características não serem consideradas para a avaliação [20, 24]

Ao longo dos últimos anos têm-se multiplicado os estudos sobre a classificação deste traçado e uma das técnicas de IA recorrente são as RNAs devido à sua capacidade de aprendizagem e generalização. Como referido em [3], as RNA tem vindo a ser utilizadas por exemplo na avaliação do estado fetal através de FFNN, cuja precisão da classificação é calculada em relação à interpretação do perito.

Por tratar-se de um sinal complexo, alguns autores optam por se debruçar sobre uma característica em particular, como, por exemplo, o reconhecimento de desacelerações [21]. Verificou, que em comparação com métodos convencionais, as RNA são capazes de reconhecer a quase totalidade das desacelerações e, conseqüentemente, não fornecem alarmes falsos.

A comparação da qualidade dos classificadores quando se utilizam diferentes arquiteturas de rede é também um objeto de estudo. O trabalho proposto em [20]

incide em diferentes classificadores com base em FFNN e *self-organizing NN* que discriminam o comportamento fetal e condições fetais normais e patológicas. Os resultados da experiência demonstraram uma elevada precisão na classificações das MLP. No entanto, indicam que a seleção das caraterísticas do sinal é extremamente importante pois influenciam o resultado da classificação, conclusões também confirmadas por outros investigadores [3, 19, 44].

A avaliação do estado fetal com base na descrição quantitativa dos sinais de FCF, bem como o desenvolvimento de sistemas automáticos para identificação de padrões normais têm sido amplamente investigados [19, 44]. O estudo descrito em [45] identifica as MLP como sendo a arquitetura mais apropriada para a avaliação do estado fetal, conclusões referidas também em [3]. Um outro estudo que comprova [18] que apresenta um sistema automático para deteção de padrões anormais permitindo aos clínicos intervir durante o trabalho de parto, aumentando as condições de saúde para o feto resultando num nascimento saudável. Apresenta uma importante ferramenta principalmente para clínicos inexperientes quando não está presente nenhum perito para consulta.

Toda a investigação que se tem vindo a efetuar neste âmbito de aplicação das RNA a traçados de CTG revela-se extremamente importante. Demonstra que são métodos flexíveis e adaptáveis ao meio em que se encontram envolvidas e devolvem, com elevada precisão, uma avaliação correta dos parâmetros que lhes são apresentados, refletindo-se numa correta classificação do traçado CTG.

3.2 Descrição do problema

A dificuldade algorítmica, e até mesmo linguística, em descrever as caraterísticas do CTG fez com que se recorresse a RNAs supervisionadas para se efetuar a correta classificação do CTG. As RNAs podem ser aplicadas quando existe um conjunto de dados exemplificativos do problema que serão transformados em vetores de entrada e saída. Antes de serem aplicadas é muito importante efetuar-se um cuidadoso pré-processamento dos dados como operações de normalização, embaralhamento e divisão dos conjuntos de dados a classificar.

Um dos processos mais importantes num problema com RNAs é a escolha da arquitetura de rede. Esta decisão não tem por base nenhuma regra específica tendo que ser escolhida aquela que se molde melhor ao tipo de problema em estudo. Da mesma forma, a definição do número de neurónios nas camadas escondidas não segue nenhum método em particular, sendo na maior parte dos casos um processo iterativo que deve contemplar um considerável intervalo de valores que estejam de acordo com a complexidade do problema. No entanto, alguns autores efetuaram experiências de forma a se encontrar um método que guiasse essa escolha. Uma aproximação deste

valor pode ser obtida pela regra da pirâmide geométrica proposta em [46],

$$n_h = \sqrt{n_{in} \times n_{out}}, \quad (3.1)$$

em que o número de neurónios na camada escondida, n_h , é igual à raiz quadrada do número de neurónios de entrada, n_{in} , vezes o número de neurónios de saída, n_{out} . O número de neurónios na camada escondida pode ainda variar de 0,5 a 2 vezes o resultado da regra da pirâmide geométrica dependendo da complexidade do problema [47]. Outros autores referidos em [48], sugerem que o valor do número de neurónios na camada escondida, admitindo uma rede de 3 camadas, deva ser 75% do número de neurónios de entrada, ou indicam que o valor a escolher deve ser entre 0,5 a 3 vezes o número de neurónios de entrada e também é proposto que se duplique esse valor até que a *performance* da rede se comece a deteriorar.

É importante referir que o teste destes métodos de cálculo do número de neurónios admitem os restantes parâmetros envolvidos constantes pois a alteração de qualquer outro parâmetro influencia o resultado da RNA e por conseguinte a escolha do melhor valor dos neurónios.

Da mesma forma, os valores iniciais para os pesos sinápticos podem ser escolhidos aleatoriamente. Apesar de existirem métodos de iniciação destes valores, esta opção é tomada por prevenir a possibilidade de alguns valores das saídas iniciais poderem tomar valores na região de saturação da função de ativação.

Nas secções seguintes são apresentadas as diferentes etapas do algoritmo de classificação: na subsecção 3.3.1 a ferramenta utilizada para desenvolvimento do algoritmo, seguindo-se as metodologias de pré-processamento dos dados, seleção da arquitetura de rede e parâmetros e, por fim, os resultados obtidos pela classificação dos dados de CTG.

3.3 Algoritmo desenvolvido

3.3.1 *Matlab*

A ferramenta utilizada para este trabalho foi o *MATLAB* (*MATrix LABoratory*), v.7.9.0(R2009b), que se trata de um *software* com um ambiente de programação para o desenvolvimento de algoritmos, análise de dados, visualização e computação numérica [49].

Mais concretamente, foi utilizada a *Neural Network ToolboxTM* que fornece ao utilizador ferramentas para a conceção, implementação, visualização e simulação de RNAs. Esta *Toolbox* é extremamente útil visto que em alguns problemas a análise formal seria difícil ou impossível, como é o caso do reconhecimento de padrões e dos sistemas de identificação e de controlo não-linear [49, 50].

3.3.2 Conjuntos de dados

Neste estudo de classificação utilizaram-se diferentes conjuntos de dados para tentar generalizar o algoritmo para qualquer problema. A tabela 3.1 contém um sumário das características dos dados utilizados, que serão descritos a seguir.

Tab. 3.1: Conjunto de dados reais utilizados neste trabalho.

Dados	#Dados	# Caraterísticas	# Classes
Iris	150	4	3
Yeast	1484	8	10
CTG	2126	16	10

Iris

O conjunto de dados Íris [51] é provavelmente o mais conhecido da literatura em reconhecimento de padrões e é um dos mais referenciados nesta área. O conjunto de dados contém 150 elementos, 4 caraterísticas e 3 classes, cada uma correspondente a um tipo de planta Íris (Íris Setosa, Íris Versicolour, Íris Virgínica). Uma classe é linearmente separável das outras duas que não são separáveis linearmente.

As caraterísticas tratam-se de valores numéricos e são:

1. Comprimento da Sépala em *cm*
2. Largura da Sépala em *cm*
3. Comprimento da Pétala em *cm*
4. Largura da Pétala em *cm*

O conjunto de dados Íris [52] possui 50 elementos em cada classe.

Yeast

Estes dados correspondem à localização de proteínas de leveduras. Pode ser encontrado em [52], possui 1484 elementos, 9 caraterísticas e existem 10 classes.

Em estudos realizados com estes dados admite-se como resultado de precisão 55% dos dados classificados. Os principais atributos do conjunto de dados de leveduras são:

1. Número de acesso à base de dados SWISS-PROT
2. Método de *McGeoch* para reconhecimento da sequência do sinal (mcg)
3. Método de *von Heijne* para reconhecimento da sequência do sinal (gvh)
4. Valor da previsão da região abrangendo a membrana ALOM (alm)

5. Valor da análise discriminativa dos aminoácidos na região N-terminal de mitocôndrias ou proteínas não mitocondriais (mit)
6. Atributo binário referente à presença de cadeias HDEL (Funcionam como sinal para a retenção no lúmen do RE) (erl)
7. Sinal de direcionamento peroxissomal no terminal-C (pox)
8. Pontuação da análise discriminante do teor de aminoácidos de proteínas vacuolares e extracelulares (vac)
9. Pontuação da análise representativa dos sinais de localização nucleares de proteínas nucleares e não nucleares (nuc)

Por sua vez, a tabela 3.2 demonstra como é efetuada a distribuição dos dados pelas respectivas classes.

Tab. 3.2: Distribuição do conjunto de dados *Yeast* pelas respectivas classes.

<i>Classe</i>	Descrição	# Elementos
CYT	Citosólico ou citoesqueleto	463
NUC	Nuclear	429
MIT	Mitocondrial	244
ME3	Proteína de membrana, nenhum sinal terminal-N	163
ME2	proteína de membrana, o sinal não clivado	51
ME1	Proteína de membrana de sinal, clivados	44
EXC	Extracelular	37
VAC	Vacuolar	30
POX	Peroxisomal	20
ERL	Lúmen do retículo endoplasmático	5

CTG

O conjunto de dados de CTG pode ser igualmente encontrado em [52]. Corresponde a 2126 CTGs processados automaticamente que foram classificados por obstetras especializados (de acordo com protocolos clínicos) e a etiquetagem de classificação de consenso foi atribuída a cada um deles. Essa classificação tanto pode ser em relação a um padrão morfológico (A, B, C ...) como a um estado fetal (N, S, P). Por conseguinte, o conjunto de dados pode ser usado tanto para experiências de 10 ou de 3 classes.

De acordo com trabalhos já efetuados [36], descartaram-se 6 das 22 características principais, sendo que as usadas foram:

1. Valor da Linha Base em bpm

2. Número de Acelerações
3. Número de Contrações Uterinas
4. Percentagem de tempo com anormal variabilidade de curto prazo
5. Valor médio da variabilidade de curto prazo
6. Percentagem de tempo com anormal variabilidade de longo prazo
7. Valor médio da variabilidade de longo prazo
8. Número de Desacelerações leves
9. Largura do histograma (histograma da FC em bpm)
10. Baixa frequência do histograma
11. Alta frequência do histograma
12. Número de picos do histograma
13. Média do histograma
14. Mediana do histograma
15. Variância do histograma
16. Tendência do histograma

A tabela 3.3 apresenta a distribuição dos dados pelas 10 classes admitidas:

Tab. 3.3: Distribuição do conjunto de dados CTG pelas respectivas classes.

<i>Classe</i>	Descrição	# Elementos
A	Sono calmo	384
B	Sono REM (<i>Rapid Eye Movement</i>)	579
C	Vigilância calma	53
S	Vigilância ativa	81
SH	Mudança de Padrão	72
AD	Padrão de acelerações/desacelerações	332
DE	Padrão Desacelerativo	252
LD	Padrão Largamente Desacelerativo	107
FS	Padrão Plano-sinusoidal	69
SUSP	Padrão Suspeito	197

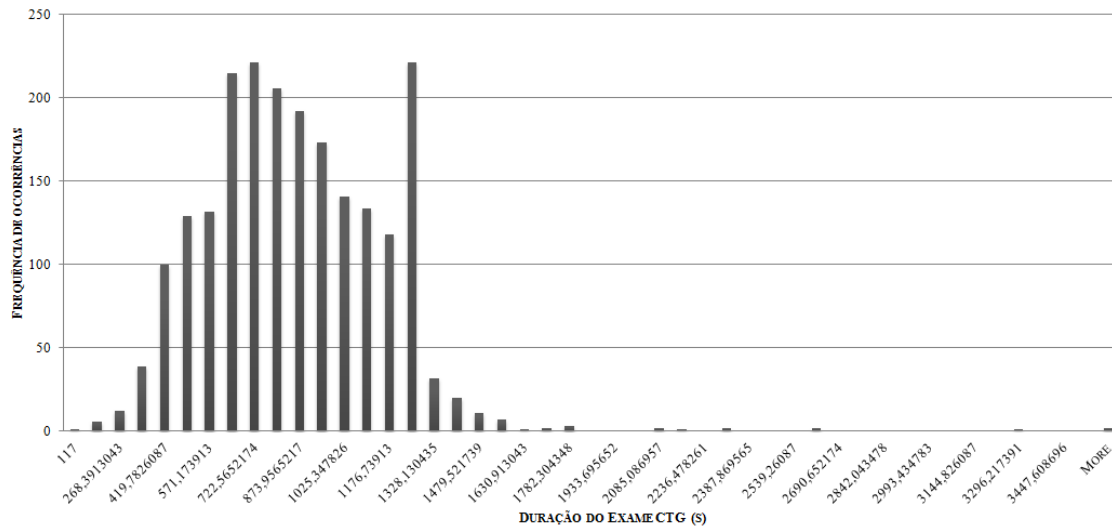
Por sua vez, a tabela 3.4 mostra a divisão dos dados pelas classes normal, suspeito e patológico.

Tab. 3.4: Distribuição do conjunto de dados CTG por 3 classes.

<i>Classe</i>	# Elementos
Normal	1655
Suspeito	295
Patológico	176

3.3.3 Pré-Processamento dos dados

A tarefa de pré-processamento dos dados representa um ponto crucial no desenvolvimento do algoritmo para que se assegure que é apresentada à RNA um conjunto de dados corretos e representativos do problema que se pretende classificar. Como se descreve a seguir, foram efetuadas diferentes operações para os dados utilizados. Relativamente aos dados de CTG, devido ao seu número de entradas e distribuição dos casos, optou-se por eliminar algumas entradas para não induzir a incorreta classificação. Como se verifica na figura 3.1, a distribuição dos dados é ampla, estando maioritariamente concentrados entre as durações médias de 250 a 1700 segundos. De forma a concentrar o conjunto de dados eliminaram-se as entradas com duração do CTG menor que 250 segundos e maior que 1700, considerando-se 2101 casos que serão melhor entradas para o treino da RNA.

**Fig. 3.1:** Número de ocorrências por diferente duração do exame CTG.

Uma outra etapa de pré-processamento dos dados está ligada com um melhor arranjo dos dados para se tentar alcançar resultados de classificação mais satisfatórios. Procedeu-se, para tal, ao embaralhamento dos dados, também conhecido como *shuffle*, que se realiza em cada etapa da aplicação do algoritmo, evitando assim um treino “viciado” numa determinada ordem dos dados. Esta tarefa, efetuada antes do treino da rede, tem um papel importante visto que esta mistura dos vetores de

forma aleatória, segundo alguns autores, melhora a precisão com que o treino é feito [53].

A divisão do conjunto de dados em subconjuntos de treino, validação e teste, é outra tarefa de extrema importância pois é necessário garantir que, ao efetuarmos esta divisão, se mantém, nos subconjuntos, a mesma distribuição dos dados relativamente às classes. Geralmente são considerados três subconjuntos: de treino, validação e teste. No entanto, alguns autores, utilizam somente os subconjuntos de treino e teste em conjunto com o método da validação cruzada (*cross validation*). Neste trabalho, foi utilizada a designada *10 fold cross validation* que consiste em dividir aleatoriamente os dados em 10 subconjuntos em que a cada experiência, 9 são utilizados para treino e 1 para teste da rede. A experiência é repetida 10 vezes de forma a que, a cada vez um dos 10 subconjuntos seja utilizado como conjunto de teste.

É importante salientar que esta operação de *cross validation* provavelmente influencia os resultados finais, isto é, para o mesmo conjunto de dados mas considerando diferentes divisões, os resultados serão consequentemente diferentes [36, 54].

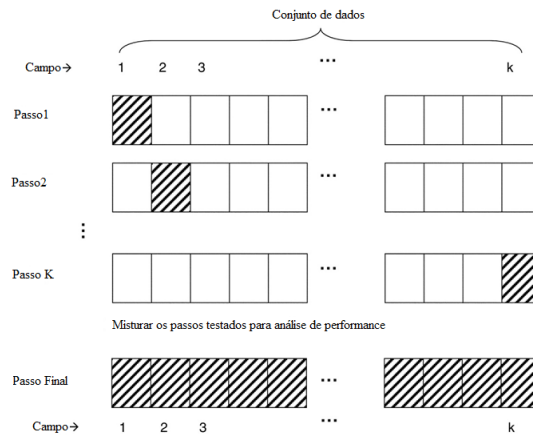


Fig. 3.2: Esquema representativo da *k-fold cross validation* [55].

Normalização

A tarefa de normalização é um dos processos mais comuns de pré-tratamento dos dados cujo objetivo é garantir que o domínio dos dados inseridos na rede vá de encontro ao domínio da função de ativação selecionada.

Na tabela 3.5 podem observar-se as fórmulas mais utilizadas para efetuar a normalização. Na notação apresentada x_n corresponde ao valor normalizado, x_0 ao valor de entrada, x_{min} , x_{max} , \bar{x} e s correspondem ao valor mínimo, máximo, média e desvio padrão respetivamente, do vetor de dados original.

Neste trabalho efetuou-se a normalização dos dados de forma a que a amplitude dos dados estivesse no intervalo $[0,1]$. Em experiências comparativas, apresentadas

Tab. 3.5: Diferentes tipos de normalização.

Intervalo [0,1]	$Xn = \frac{x_0 - x_{min}}{x_{max} - x_{min}}.$
Intervalo [a,b]	$Xn = (b - a) \frac{x_0 - x_{min}}{x_{max} - x_{min}} + a .$
Estatística	$Xn = \frac{(x_0 - \bar{x})}{s}.$
Simples	$Xn = \frac{x_0}{x_{max}} .$

em anexo A.2, utilizaram-se outras funções de normalização de forma a perceber a influência nos resultados.

3.3.4 Metodologia de Classificação

O algoritmo de classificação utilizado teve por base uma MLP com apenas uma camada escondida que, pela literatura, será o suficiente para o nosso caso de estudo [34].

Neste trabalho adotou-se para todos os casos de classificação a FFNN através da utilização da função *newff* que requer como parâmetros de entrada os vetores de entrada e respetivos alvos (X e T), o número de neurónios na camada escondida e de saída (n_h e n_{out}), as funções de ativação das camadas escondida e de saída, o algoritmo de treino, a função de *performance* e de divisão dos dados. Definem-se também valores para o número máximo de épocas, gradiente mínimo pretendido e taxa de aprendizagem. Nestas experiências, a função de ativação utilizada nas camadas escondida e de saída foi a *logsig*, de acordo com a amplitude de saída dos dados normalizados.

Iterativamente foram selecionados os melhores parâmetros para a execução da rede, como o número de neurónios para a camada escondida e o valor da taxa de aprendizagem. Para a divisão dos dados na RNA foi utilizada a função *dividerand* que forma 3 conjuntos utilizando índices aleatórios. Por defeito, a proporção padrão para o treino é:

```
net.divideParam.trainRatio=0.6;
net.divideParam.valRatio=0.2;
net.divideParam.testRatio=0.2;
```

No entanto, visto não se considerar conjunto de dados para validação, a proporção utilizada pelo algoritmo foi definida após o resultado de várias experiências para determinação dos melhores parâmetros.

Os algoritmos de treino utilizados tiveram por base várias experiências de maneira a verificar qual o que apresenta melhores resultados. Inicialmente testou-se com a função *trainlm* que corresponde ao algoritmo de *Levenberg-Marquardt*. Devido aos resultados obtidos com os dados auxiliares, apresentados no Anexo A.1, este algoritmo não se revelou o mais indicado necessitando de maior tempo de execução. Desta feita, admitiram-se as funções de treino: *traingd* e *traingda* que correspondem à atualização de pesos da rede de acordo com a descida do gradiente, sendo que esta última admite também uma taxa de aprendizagem adaptativa retrógrada [50].

Uma vez criada, a RNA é treinada com recurso à função *train* que necessita da informação de que rede deve ser treinada bem como os respetivos conjuntos de treino (X_{treino} e T_{treino}). O treino é então efetuado até se atingir um dos critérios de paragem podendo ser o número máximo de épocas alcançado, máximo de tempo excedido, o gradiente de desempenho ter um valor abaixo do parâmetro *net.trainParam.min_grad* fornecido, sendo que estes foram já referidos no capítulo anterior.

Na figura 3.3 pode consultar-se a janela de visualização desta etapa que se trata de uma aplicação que informa o utilizador do modo como a rede está a ser treinada, podendo este verificar se vai ou não de encontro aos parâmetros inicialmente introduzidos.

Seguidamente realiza-se a simulação da rede, desta vez recorrendo aos conjuntos de dados de teste para poder verificar se a rede aprendeu ou não corretamente. O resultado desta operação necessita de um pós-processamento, ou seja, colocar os resultados obtidos num formato em que se consiga efetuar a correta comparação com as saídas que se pretendem. É então realizada a transformação da saída da simulação Y para o modo de classes, sendo que estas são definidas de forma binária¹.

Após esta fase, é necessário uma forma de avaliar os erros de classificação. Adotou-se como medida de *performance* de classificadores a exatidão (*accuracy*), recorrendo à utilização das matrizes de confusão.

O cálculo dos erros baseou-se então na diferença entre a saída da simulação da rede e os valores desejados. O algoritmo foi repetido 10 e 20 vezes sendo o resultado final a média e o desvio padrão do conjunto das experiências. Esta operação dá-nos informação sobre a taxa de erros existentes no algoritmo. Para podermos obter informação sobre a exatidão da classificação, ou seja, da percentagem de casos corretamente classificados, efetua-se o complemento da taxa de erro (1- taxa de erro).

¹ A codificação binária é um sistema de numeração baseado em apenas dois algarismos: 0 e 1.

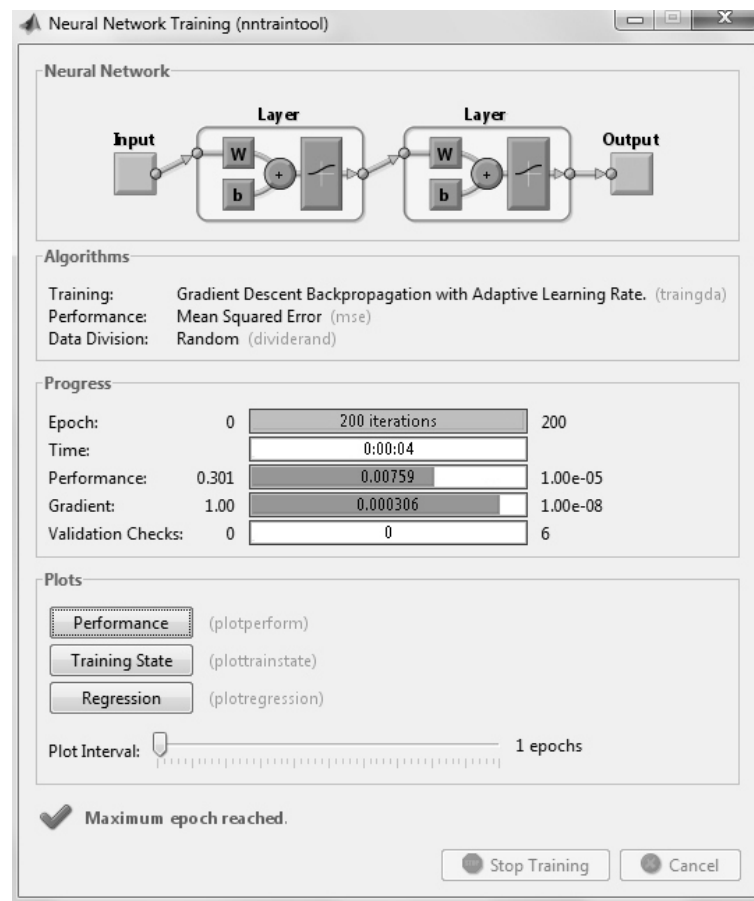


Fig. 3.3: Janela de visualização do treino da RNA.

Uma das técnicas também utilizadas para se obter mais informação sobre o desempenho do algoritmo é a matriz de confusão que se identificou, para cada iteração, e indica o número de classificações corretas em relação ao conjunto de alvos pretendidos [56]. Na figura 3.4 pode consultar-se um exemplo de uma matriz de confusão numa experiência dos dados Íris em que as colunas representam os casos reais e as linhas os casos previstos pelo algoritmo. O número de acertos em cada caso está indicado na diagonal principal da matriz.

		Valor Real		
Valor Previsto	50	0	0	
	0	26	5	
	0	24	45	

Fig. 3.4: Janela de visualização da matriz de confusão para os dados Íris.

3.4 Classificação dos conjuntos de dados

3.4.1 Resultados dos conjuntos de dados auxiliares

Para este estudo utilizaram-se como descrito outros conjuntos de dados de forma a generalizar o algoritmo e também ver se este funciona em diferentes situações de classificação. Nestes casos, considerou-se apenas a partição do conjunto de dados em 70% para treino e 30% para teste. Os dados foram normalizados para o intervalo $[0,1]$ selecionando a mesma função de ativação para as camadas escondida e de saída, neste caso a *logsig*.

Para o conjunto de dados Íris efetuaram-se experiências para determinar apenas qual o melhor número de neurónios para as camadas escondidas e admitiu-se o número na saída igual ao número de classes sendo neste caso 3. No Anexo A.1.2 encontram-se os respetivos valores do MSE para os diferentes números de neurónios. Escolheu-se aquele que apresentava menor erro sendo de 10 neurónios para o treino com *traingd* e de 4 para *traingda*. As MLPs criadas com base nestes parâmetros efetuaram o treino durante 200 épocas, com uma taxa de aprendizagem de 0,01. Obteve-se como resultado da classificação os erros (média e desvio padrão) apresentados na tabela 3.6, para diferentes valores de N .

Tab. 3.6: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados Íris usando a normalização no intervalo $[0,1]$.

N	<i>traingda</i>	<i>traingd</i>
10	$0,1057 \pm 0,1383$	$0,6593 \pm 0,1382$
20	$0,1397 \pm 0,1753$	$0,6700 \pm 0,1482$

Tab. 3.7: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados Íris usando a normalização no intervalo $[-1,1]$.

N	<i>traingda</i>	<i>traingd</i>
10	$0,2850 \pm 0,1509$	$0,5693 \pm 0,1749$
20	$0,0760 \pm 0,1078$	$0,5010 \pm 0,1904$

Pela consulta das tabelas 3.6 e 3.7 verifica-se que o melhor resultado de classificação obtido foi com a MLP com 4 neurónios na camada escondida, com a função *traingda* e função de ativação *tansig*. O erro foi de 0,0760, sendo portanto a percentagem de classificações corretas de 92,4%.

Relativamente aos dados *Yeast* efetuaram-se o mesmo conjunto de experiências anteriores, estando também no Anexo A.1.2 os resultados para os melhores parâmetros para treino da rede. Neste caso, o número de neurónios da camada escondida é

de 30, com uma taxa de aprendizagem de 0,05 e 0,01 para os respetivos algoritmos de treino.

Tab. 3.8: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados *Yeast* usando a normalização no intervalo $[0,1]$.

N	<i>traingda</i>	<i>traingd</i>
10	$0,6827 \pm 0,0872$	$0,8791 \pm 0,1176$
20	$0,4791 \pm 0,0513$	$0,8889 \pm 0,1117$

Tab. 3.9: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados *Yeast* usando a normalização no intervalo $[-1,1]$.

N	<i>traingda</i>	<i>traingd</i>
10	$0,4669 \pm 0,0686$	$0,6523 \pm 0,0865$
20	$0,4741 \pm 0,0512$	$0,8286 \pm 0,1208$

Pela consulta das tabelas 3.8 e 3.9 verifica-se que a melhor classificação obtida para o conjunto de dados *Yeast* foi uma média de erros de 46,69 correspondendo a uma percentagem exata de classificações de 53%. Este resultado foi obtido para a MLP com função de ativação *tansig* para 10 iterações do algoritmo.

3.4.2 Resultados da classificação do CTG

Na classificação dos dados dos traçados de CTG teve-se em consideração as principais características que permitem definir o estado fetal. Admitiu-se uma classificação mais completa de 10 estados fetais possíveis, representando estes cada uma das classes assumidas no algoritmo. Com esta classificação é possível obter-se melhor informação do que se se considerasse apenas 3 classificações (Normal, Suspeito e Patológico), no entanto, a utilização das 10 classes torna o problema bastante mais complexo já que este é um conjunto não balanceado no que diz respeito ao número de elementos por classe.

Primeiramente foi necessário descobrir que parâmetros seriam os melhores para a aprendizagem da rede. Foi criada uma FFNN onde se variou o número de neurónios nas camadas escondidas entre 10 e 50 e a taxa de aprendizagem com valores de 0,01 e 0,05. A seleção dos melhores valores foi feita com base na *performance* do treino desta rede, neste caso a função MSE, selecionando-se aqueles cujo valor fosse o mínimo.

Teve-se em atenção a forma como a distribuição dos dados é apresentada à rede. Como referido, não se considerou um conjunto de dados específico para validação.

Contudo, a variação das percentagens de dados para treino e teste foram variadas apenas para o conjunto de dados de CTG nesta primeira etapa de definição dos melhores parâmetros. Desta feita, admitindo a sequência de treino-validação-teste, as distribuições utilizadas foram 60%-0-40%, 70%-0-30% e 90%-0-10%. Alternaram-se também os algoritmos de treino de forma a constatar a sua influência na aprendizagem da rede. As tabelas 3.10, 3.11 e 3.12 apresentam os resultados obtidos para o MSE nas diferentes considerações do parâmetro η , que representa a taxa de aprendizagem, e do número de neurónios, representado por n_h , em cada experiência realizada.

Tab. 3.10: Resultados de *performance* da rede com 60% para treino e 40% para teste.

η	nº neurónios	traingda	traingd
0,01	10	0,1019	0,4017
	20	0,0098	0,2049
	30	0,0283	0,2124
	40	0,0458	0,0138
	50	0,0392	0,1112
0,05	10	0,2033	0,1022
	20	0,1071	0,0093
	30	0,0392	0,0097
	40	0,0566	0,2132
	50	0,0363	0,0190

Tab. 3.11: Resultados de *performance* da rede com 70% para treino e 30% para teste.

η	nº neurónios	traingda	traingd
0,01	10	0,5015	0,2021
	20	0,2149	0,1168
	30	0,0195	0,0087
	40	0,1137	0,0122
	50	0,0182	0,0148
0,05	10	0,2023	0,2024
	20	0,2038	0,3106
	30	0,1100	0,3009
	40	0,0058	0,0100
	50	0,0754	0,0203

Pelos resultados de *performance* obtidos verifica-se que a melhor divisão dos dados é a de 70% dos dados para treino e 30% teste, para a função *traingda*, em que os melhores parâmetros foram de 40 e 0,05 para o número de neurónios e taxa de aprendizagem, respetivamente. Por sua vez, em relação à função *traingd* esses valores são de 30 e 0,05. Tendo por base estes valores criaram-se duas MLP: para

Tab. 3.12: Resultados de *performance* da rede com 90% para treino e 10% para teste.

η	nº neurónios	traingda	traingd
0,01	10	0,2692	0,1015
	20	0,1201	0,1938
	30	0,0337	0,1089
	40	0,0424	0,0152
	50	0,0483	0,1150
0,05	10	0,3072	0,1032
	20	0,2277	0,2121
	30	0,0438	0,1047
	40	0,0409	0,0152
	50	0,0381	0,1149

o treino com o algoritmo da *traingda* a melhor arquitetura apresentada é de [40:10] com $\eta = 0,05$. Para o outro algoritmo de treino, a MLP apresenta a arquitetura de [30:10] e $\eta = 0,01$. Em ambos os casos o número máximo de épocas foi fixado em 500.

Tab. 3.13: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG usando a normalização no intervalo [0,1].

N	<i>traingda</i>	<i>traingd</i>
10	0,3566 \pm 0,0428	0,9005 \pm 0,0691
20	0,3570 \pm 0,0393	0,8865 \pm 0,0749

Na tabela 3.13 encontram-se os respetivos erros de classificação relativos ao conjunto de dados CTG comprovando que o melhor resultado da rede foi para 10 iterações com a função *traingda*. A MLP obteve assim uma taxa média de erro equivalente 35,66% sendo a percentagem de classificações corretas de 64,34%.

Na figura 3.5 apresenta-se um exemplo de uma matriz de confusão para este conjunto de dados, de uma iteração da RN. Ao longo do treino da rede são cometidos vários erros de classificação sendo que, os valores na diagonal da matriz representam o número de casos que a rede classificou corretamente, nesta iteração do algoritmo.

Adicionalmente, para verificar se a classificação dos dados CTG obtém melhores resultados, procedeu-se à alteração das funções de ativação das camadas de entrada e saída da rede para a *tansig*. Esta alteração implicou a normalização dos dados de entrada da RN para o intervalo [-1,1]. Os restantes parâmetros da rede foram mantidos, ou seja, apenas se efetuou o mesmo número de experiências com os parâmetros que obtiveram menor valor de erro.

Verifica-se, pela análise da tabela 3.14 que, apesar da alteração das funções de ativação, o melhor resultado obtido continua a ser com o algoritmo de treino *traingda*.

```

confusion_matrix_test(:, :, 6) =

    29     1     1     0     1     0     0     0     1     4
     0    49     1     4     0     2     0     0     0     1
     0     0     4     0     0     0     0     0     0     0
     0     1     0     7     0     0     0     0     0     0
     0     0     0     0     6     0     0     0     1     0
     0     1     0     0     0    27     2     0     0     2
     1     0     0     0     0     0    23     1     0     0
     0     0     0     0     0     0     0    10     0     0
     0     0     0     0     0     0     0     0     3     3
     1     1     0     0     2     0     0     0     3    12

```

Fig. 3.5: Matriz de Confusão obtida para uma iteração da RNA.

Tab. 3.14: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG usando a normalização no intervalo $[-1,1]$.

N	<i>traingda</i>	<i>traingd</i>
10	$0,3812 \pm 0,0448$	$0,8622 \pm 0,0779$
20	$0,3742 \pm 0,0471$	$0,8632 \pm 0,0739$

A RN classifica 62,55% dos casos corretamente, tendo um erro médio de 0,3742.

Pode concluir-se que, apesar dos resultados de escolha de parâmetros, o algoritmo de treino que considera a descida do gradiente com taxa de aprendizagem adaptativa se revela melhor para este caso particular de classificação. Combinando as funções de transferência da camada escondida e da camada de saída, concluiu-se que a combinação *logsig/logsig* tem, neste caso, maior eficácia que a combinação *tansig/tansig*.

3.5 Conclusões

A classificação de CTG é um problema complexo devido não só ao elevado número de dados mas também às muitas características que é necessário ter em conta e posteriormente à codificação das diferentes classes.

É preciso recolher e preparar um grande volume de dados, tendo o cuidado de verificar as condições de preenchimento das variáveis e, caso necessário, eliminar registos sobre os quais se tem algumas dúvidas e consequentemente que possam influenciar o resultado da classificação.

Geraram-se conjuntos aleatórios de 60%-0%-40%, 70%-0%-30% e 90%-0%-10% (identificando cada percentagem a dimensão do conjunto de treino, validação e teste respetivamente), a fim de conferir diferentes dimensões de parametrização às redes em comparação e garantir a capacidade de generalização que é essencial neste tipo de trabalho.

Os resultados apresentados ajudaram a perceber que, mesmo tratando-se de um

conjunto de amostras pequeno, a RNA aprende e comporta-se de maneira distinta para diferentes funções de transferência. Conseguiu-se também entender que para diferentes tipos de treino e suas possíveis variações como o número de nós na camada escondida ou o valor alvo para o erro, a rede consegue aprender.

Comparativamente aos trabalhos já existentes na área verifica-se que os resultados obtidos não apresentam grande variação, sendo portanto aceitáveis as percentagens de classificações quer dos conjuntos de dados auxiliares bem como no caso dos dados dos CTGs. No entanto, não foi utilizado em todos os estudos as mesmas características, parâmetros e arquiteturas de rede sendo que estes factos influenciam os resultados que se obtêm. Como se refere na literatura, é necessário existirem regras bem definidas para homogeneizar as classificações e facilitar a interpretação dos resultados. O razoável é então utilizar as RNAs como complementos dos métodos convencionais utilizados requerendo um alto nível de confiabilidade nos resultados [19, 20, 21].

Previsão do sinal CTG

O presente capítulo foca-se na abordagem da previsão temporal do sinal de CTG através de RNAs. Inicialmente apresentam-se diferentes estudos de previsão de séries temporais e a sua importância. É apresentado o principal objetivo deste capítulo e feita uma breve descrição das séries temporais usadas e como o conjunto de dados a origina. Posteriormente são apresentados os resultados para a previsão tendo por base o algoritmo criado.

4.1 Estado da arte

A previsão de eventos futuros é uma tarefa importante e as RNAs têm sido amplamente utilizadas neste tipo de problemas por se revelarem como alternativas às técnicas tradicionais: apresentam capacidades de generalização e aproximação que permitem um melhor desempenho de previsão.

Inúmeras áreas têm vindo a aplicar RNAs para efetuar previsões financeiras, turísticas, meteorológicas, de tráfego, de ações de bolsa ou de modelação de séries biomédicas, entre muitas outras. Em resumo, em situações cujos eventos futuros necessitam de ser previstos com base em informação ocorrida no presente e passado.

Na maior parte dos casos a previsão é baseada numa série temporal, ou seja, consiste num número de observações de um determinado evento durante um período de tempo e requer a previsão de um evento particular para um tempo específico (dia, hora, segundo). As séries temporais podem considerar-se lineares ou não lineares consoante o resultado a prever esteja linear ou não-linearmente relacionado com o corrente resultado [15, 57, 58].

As primeiras aplicações de RNA para previsão remontam aos anos 80 e desde então começaram-se a aplicar nas mais diversas áreas [17].

A área do turismo é um exemplo de aplicação destas técnicas e muitos trabalhos se têm desenvolvido com o objetivo, por exemplo, de captar as tendências turísticas

de determinada região ou prever o número de dormidas num determinado hotel [59, 60, 61, 62]. Na figura 4.1 pode ver-se um exemplo de uma série temporal da área do turismo, que representa o número de dormidas no intervalo de Janeiro de 1987 a Dezembro de 2009.

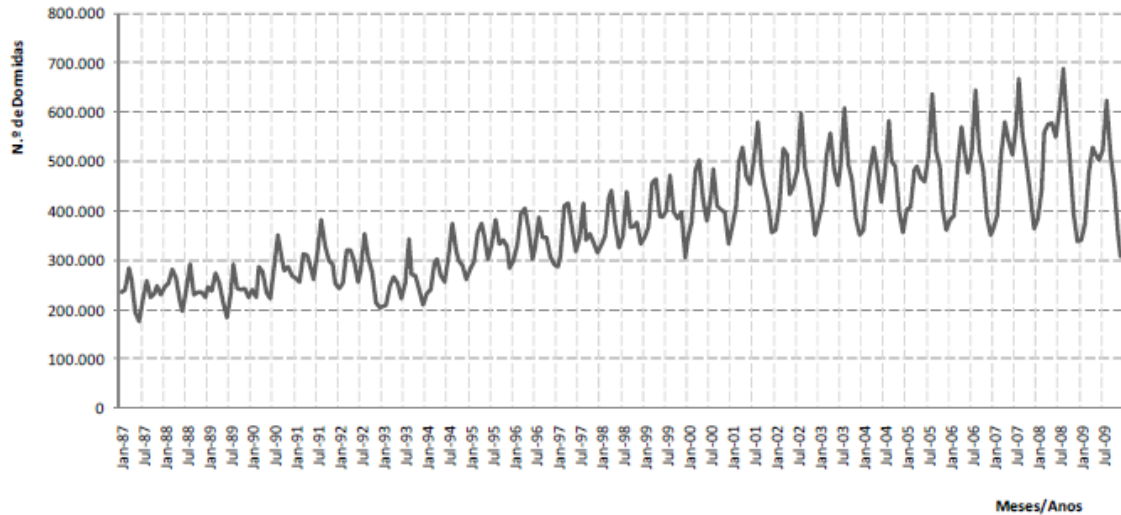


Fig. 4.1: Dormidas dos turistas nas Unidades de Alojamento da Região da Madeira, no intervalo [Jan-87:Dez-09] [62].

A temperatura, radiação e estado do tempo são também pontos de interesse de muitos autores. Alguns autores estudaram a previsão do tempo de 1 a 10 dias tendo em conta as inúmeras variáveis que este tipo de problemas possui e comprovaram que o uso de RNA nestes casos é bastante favorável [63]. Outros investigaram o potencial do uso de RNA para desenvolver modelos para a previsão da radiação solar global em determinadas zonas como *Al ain* nos Emirados Árabes Unidos [64] e na China [65].

A previsão de eletricidade, por exemplo, tem sido uma parte essencial de um planeamento eficiente de energia. O objetivo da previsão de carga a curto prazo é de prever demandas futuras de energia elétrica com base em dados históricos e outras informações, como a temperatura [57, 66, 67, 68]. A ajuda da previsão com RNA pode fazer-se no planeamento de vendas e distribuição [17], na regulação de temperaturas, na previsão de fluxos de águas [64] bem como no setor financeiro para prever desde a variação da taxa de juros, preços de ações entre muitos outros cenários [69, 70].

Na literatura são referidas muitas áreas de aplicação desta técnica de IA na previsão de séries temporais. Contudo, não existe nenhuma regra ou método que decida qual a melhor arquitetura de rede a adotar para casos de previsão. Obviamente é sempre necessário ter-se em conta o conjunto de dados que se analisa e consequentemente o resultado que se pretenderá obter [71].

Existem diferentes arquiteturas de rede na aplicação da previsão com RNA referenciadas pela literatura como por exemplo as redes *Elman*, NARX (NonLinear Autoregressive model with Exogenous Inputs), RRBFN¹ ou *Jordan* [34, 58, 66, 67, 72, 73].

Na área médica os estudos efetuados baseiam-se na previsão da incidência de uma determinada doença, como [74] que estudou o comportamento de RNAs na previsão da incidência da dengue no Brasil. Surgem também estudos relacionados com a previsão do eletroencefalograma (EEG) [75] onde os autores utilizam redes recorrentes Elman para prever o sinal um único passo à frente. No entanto, relativamente ao sinal de CTG não se encontram muitos trabalhos sobre a previsão do sinal. Em [8], os autores tentam prever uma determinada característica, mais precisamente a acidemia neonatal, com base nas características do sinal CTG.

4.2 Descrição do problema

O ideal em qualquer sinal fisiológico seria conseguir prever-se instantes futuros tendo por base os parâmetros mais importantes do sinal em instantes passados, contudo esta situação não acontece na prática.

O principal objetivo deste capítulo é aplicar um algoritmo de previsão à série temporal correspondente ao sinal de CTG. A importância desta tarefa advém do facto de se conseguir perceber o estado fetal antecipadamente, ou seja, prever n instantes de tempo à frente permitiria obter mais informações sobre as condições em que estaria o feto, nomeadamente se poderia existir ou não indício de algum tipo de anomalia. A grande vantagem será obviamente a possibilidade de uma resposta por antecipação evitando graves problemas para o feto.

Pode considerar-se a previsão de séries temporais como sendo um problema de processamento de sinal em que, a partir de uma sequência de amostras N de uma determinada variável, $x(n), x(n-1), \dots, x(n-N+1)$, uniformemente espaçadas no tempo, se pretende obter uma estimativa para o próximo elemento da série, $\hat{x}(n+1)$.

Este tipo de previsão é designado por um passo adiante (UPA) em que apenas se prevê o próximo valor da série temporal, sem que haja realimentação do valor previsto para a entrada da rede, como se comprova pela figura 4.2 (a). Por sua vez, pode preferir-se efetuar uma previsão com um horizonte mais amplo, isto é, múltiplos passos adiante (MPA), em que existe realimentação da saída do modelo para a entrada, como demonstra a figura 4.2 (b). Como facilmente se percebe, a previsão MPA é mais complexa do que a UPA [16, 34].

A escolha da arquitetura de rede, como no caso da classificação, não obedece a nenhuma regra específica, não sendo por isso uma tarefa fácil. Normalmente recorre-se

¹ *Recurrent Radial Basis Function Network*

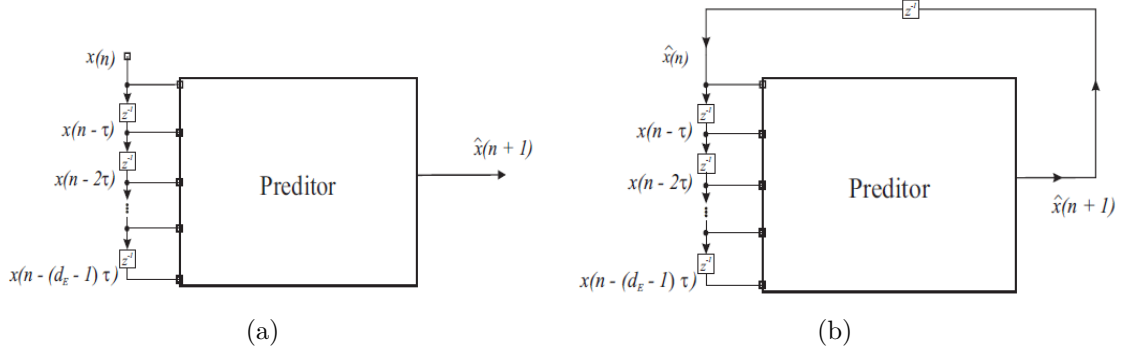


Fig. 4.2: Esquema de Previsores: (a) Sem realimentação; (b) Recursivo, com alimentação [16].

a procedimentos do tipo tentativa-erro em que apenas se simula um número limitado de cenários, tendo em consideração diferentes arquiteturas de rede, funções de ativação, preparação dos dados, algoritmos de treino, tamanho das séries utilizadas e medidas de *performance* da rede. Este trabalho focar-se-á apenas na aplicação de algoritmos de previsão baseados em redes FTDNN.

4.3 Algoritmo de previsão com base na FTDNN

4.3.1 Séries de referência

Os exemplos utilizados para previsão compreendem séries temporais caóticas, simuladas a partir de equações diferenciais e séries temporais caóticas reais. Seguidamente são apresentadas as séries de *Mackey-Glass*, de laser e a série principal deste capítulo, o sinal CTG.

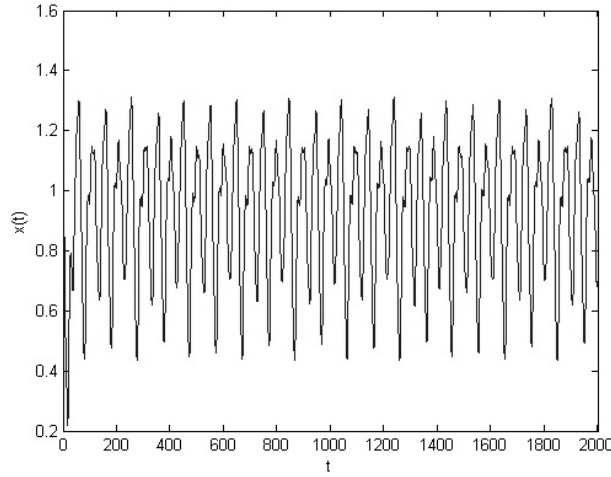
Série Caótica de *Mackey-Glass*

As observações da série temporal de *Mackey-Glass* são obtidas através da equação diferencial que se define por [76]:

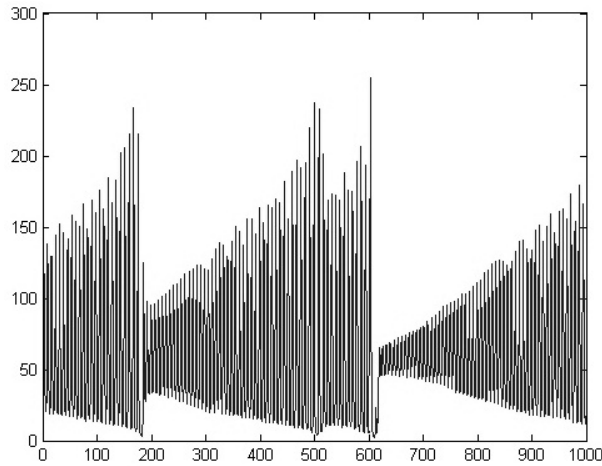
$$\frac{dx(t)}{dt} = \frac{\alpha x(t - \tau)}{1 + x(t - \tau)^{10}} - \beta x(t), \quad (4.1)$$

em que $x(t)$ é o valor da série temporal no instante de tempo t e τ o atraso introduzido na rede. Considerando $x(0) \in [0, \tau]$, o sistema converge para um ponto de equilíbrio estável se $\tau < [4, 53 - 13, 3]$ tornando-se caótico para $\tau > 16.8$, após uma série de duplicações de períodos para $\tau \in [13, 3 - 16, 8]$. Para as simulações deste trabalho, admitiram-se os valores encontrados na literatura para os seguintes parâmetros: $\alpha = 0, 2$, $\beta = 0, 1$, $\tau = 17$ e $x(0) = 1, 2$.

A equação 4.1 traduz um modelo de produção de glóbulos brancos no corpo humano e considera-se uma referência por indicar uma série temporal caótica representando as oscilações não-lineares semelhantes em muitos processos fisiológicos [16, 70, 77, 78]. Na figura 4.3 (a) está representada a série de *Mackey-Glass* com os parâmetros referidos anteriormente. As observações geradas a partir dos valores iniciais foram obtidas utilizando o método de 4ª ordem de *Runge-Kutta* [78].



(a)



(b)

Fig. 4.3: Séries de Referência: (a) Série Caótica de *Mackey-Glass*; (b) Série do Laser Caótico.

Série Caótica de Laser

O sinal caótico de laser tem origem numa sequência de medições de intensidade de pulsação de um laser NH_3 infravermelho. Inicialmente esta série foi disponibilizada como parte de uma competição de previsão de séries temporais promovida em 1992 pelo Instituto Santa Fé, nos Estados Unidos da América, cujo objetivo principal era usar os 1000 primeiros pontos da série para prever os 100 valores seguintes [79]. Na

figura 4.3 (b) está representada a série do laser caótico e percebe-se que quando um valor crítico é atingido, a série fica instável e a oscilação recomeça mas com um valor de amplitude mais baixo. Devido a esta instabilidade parte da informação da fase de oscilação é perdida e a amplitude apresenta um comportamento imprevisível.

É importante salientar que a situação crítica do laser caótico ocorre por volta do instante de tempo 200, quando ocorrem colapsos da intensidade do laser repentinamente (passam de um valor alto para um valor baixo), para então começar uma recuperação gradual da intensidade do laser [16].

Sinal CTG

O CTG, descrito anteriormente no Capítulo 3, é composto por dois sinais principais: sinal da FCF e das UC, adquiridos durante um determinado intervalo de tempo. Os dados para a construção da série temporal, referente à FCF, utilizados neste trabalho são obtidos através da versão 3.6.195 de *OmniView Cliente*.

O traçado é adquirido em milisegundos sendo as amostras normalmente 4 por segundo, exceto em situações de perda de sinal em que esse intervalo de tempo pode ser maior. O facto de existir esta perda de sinal requer que seja necessariamente efetuado um pré-processamento dos dados por forma a eliminar estas quebras no sinal. A figura 4.4 apresenta a série temporal utilizada neste trabalho.

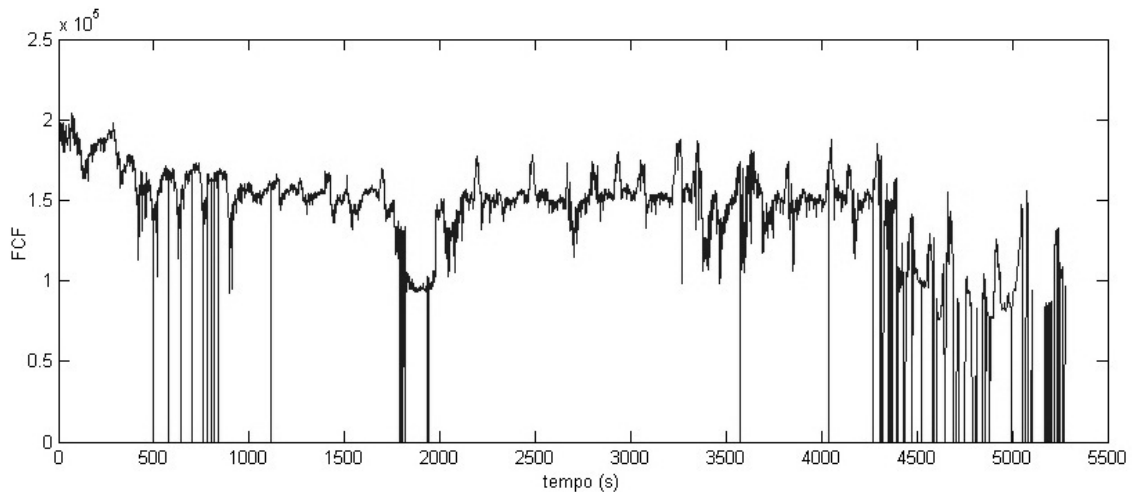


Fig. 4.4: Série temporal referente à FCF de um traçado CTG.

4.3.2 Pré-Processamento dos dados

A qualidade e o resultado da previsão estão ainda muito relacionados com a qualidade do pré-processamento do sinal. Múltiplos fatores podem influenciar a aquisição do sinal, desde a má conexão dos elétrodos a falhas com o equipamento. O sinal CTG é suscetível a este tipo de falhas e é necessário existirem operações

de pré-processamento para apresentar ao método de previsão, neste caso à RNA, o melhor conjunto de dados de forma a se garantir a obtenção dos melhores resultados possíveis.

Estudos efetuados no âmbito do CTG revelam o desenvolvimento de um algoritmo de pré-processamento para se lidar com o problema do ruído e artefactos causados por movimentos fetais e maternos. O algoritmo deteta no sinal da FCF padrões em que a linha de base esteja abaixo dos 60 bpm e cuja diferença entre picos seja superior a 25 bpm e elimina esses valores na série através da interpolação (quando os períodos de perda de sinal não excedem os 2 segundos). No que respeita a segmentos mais longos, o segmento anterior de igual comprimento é replicado ocupando esses valores em falta na série [80, 81, 82].

A escala de tempo fornecida diz respeito à duração do parâmetro, normalmente espaçados de 250 em 250 *ms*. Convertem-se a escala de tempo para segundos e colocaram-se os instantes de tempo de forma crescente de forma a facilitar a interpretação do parâmetro em relação ao tempo.

Como se verifica na figura 4.4 a representação gráfica da série apresenta inúmeros picos que dizem respeito a perdas de sinal durante a aquisição. Estes valores são comprometedores para os resultados da previsão que se pretende efetuar. De forma simplificada e rápida, a solução encontrada para eliminar estes picos foi suavizar o sinal representado através da utilização de filtros, por exemplo uma média deslizante. Desta feita, é encontrada a média de execução dos elementos do vetor de entrada X utilizando um tamanho de janela M , parâmetro escolhido por tentativas, verificando o que apresentava melhor suavização.

Na figura 4.5 mostram-se comparações da série original com o resultado após ser aplicado o filtro de suavização. Pode verificar-se que quanto maior o tamanho da janela menos ruído o sinal tem. Isto vai-se verificando através da eliminação dos picos com valor igual ou próximo de 0. Contudo, se o valor introduzido for demasiado alto, as características básicas do sinal começam por ficar comprometidas. Desta feita, adotou-se como resultado para aplicação na previsão a série correspondente ao sinal filtrado com $M=200$.

Após a filtragem do sinal efetua-se a operação de normalização igualmente decrita no capítulo 3 em que as amplitudes normalizadas das séries se encontram entre $[-1,1]$.

4.3.3 Criação do modelo de previsão

No algoritmo de previsão utilizado, após as etapas de pré-processamento, procedeu-se à divisão dos dados da série em conjuntos de treino e teste. Admitiram-se 80% para treino e os restantes 20% para teste. Também na previsão não se considerou um conjunto específico para a validação da rede. Na figura 4.6 pode verificar-se a

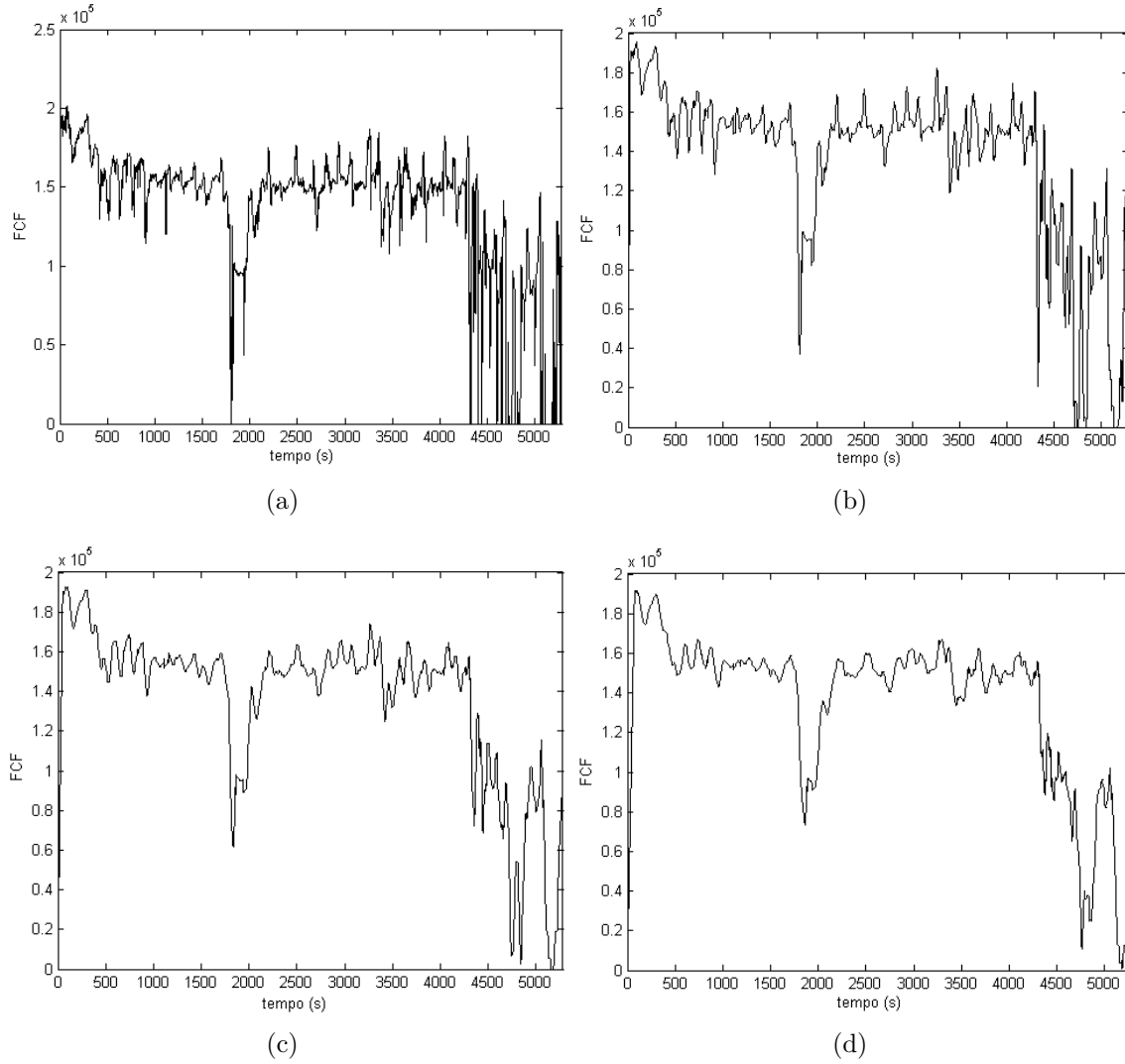


Fig. 4.5: Resultados da suavização do sinal de FCF do CTG: (a) $M=25$; (b) $M=100$; (c) $M=200$; (d) $M=300$.

distinção na série correspondente à FCF do CTG dos conjuntos de treino e teste.

A arquitetura de rede selecionada foi a FTDNN, com uma camada escondida, que exige que sejam definidos os atrasos que se pretendem induzir à rede. Admitiu-se um atraso de 100 valores da série para os diferentes passos a prever e inicialmente apenas uma camada escondida com 50 neurónios e a de saída com o número correspondente ao valor previsto pela rede. A função de ativação selecionada para ambas as camadas foi a *tansig* e admitiu-se o algoritmo de treino definido pela função *traingda*.

Como forma de avaliar o desempenho da tarefa de previsão, define-se como erro de previsão a diferença entre o valor realmente observado para a próxima amostra da série e aquele que é estimado, $\hat{x} + 1$, ou seja,

$$e = x(n) - \hat{x}(n). \quad (4.2)$$

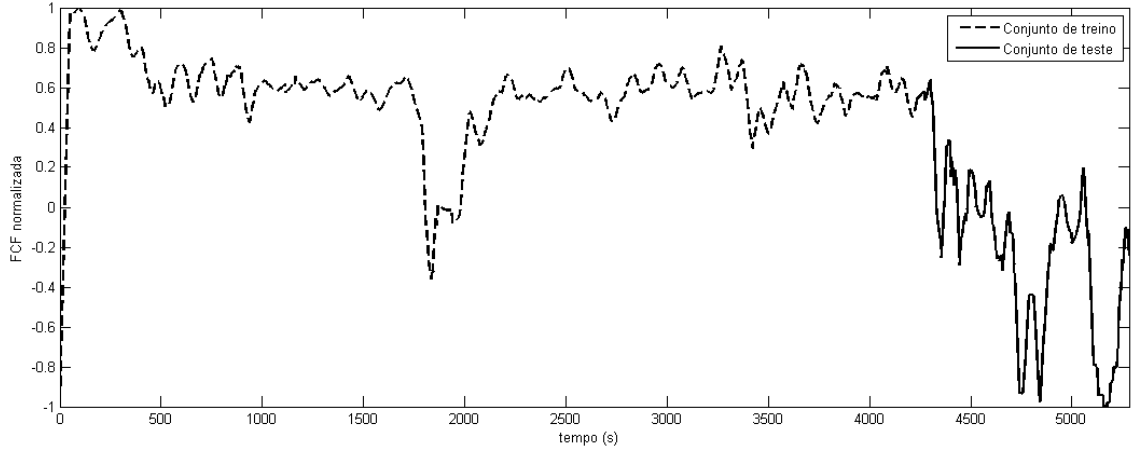


Fig. 4.6: Série CTG modificada com distinção dos conjuntos de treino e teste.

A sequência de erros $e(n)$ para $n = 1, \dots, N$ é utilizada para avaliar a precisão do modelo através do Erro Quadrático Médio Normalizado (*Normalized Mean-Squared Error*, NMSE), dado pela expressão:

$$NMSE(N) = \frac{1}{K \cdot \sigma_x^2} \sum_{n=1}^K e^2(n) = \frac{\hat{\sigma}_e^2}{\sigma_x^2}, \quad (4.3)$$

em que σ_x representa a série a ser prevista, $\hat{\sigma}_e^2$ é a variância dos erros e K é o tamanho do vetor de erros. Adotou-se esta função de *performance* pois a literatura refere que, na lógica deste estimador, na impossibilidade de gerar uma previsão mais exata de uma grandeza, adota-se o seu valor médio. Esta estratégia é utilizada por exemplo pelas companhias de energia elétrica ou água nas medições de leituras [16, 83, 84].

4.4 Resultados da Previsão

4.4.1 Resultados da previsão da série de *Mackey-Glass*

Para uma primeira avaliação do desempenho das redes neurais dinâmicas na tarefa de previsão e modelação de séries temporais caóticas a rede é treinada com a série de *Mackey-Glass*. A RN utiliza as seguintes configurações: treino pelo algoritmo de *backpropagation* com 500 épocas, taxa de aprendizagem igual a 0,01 e atraso inicial $\tau = 100$.

A série de Mackey-Glass possui 2000 pontos, sendo que os 1400 primeiros são destinadas para o treino e os restantes 600 para teste da RNA. Os resultados de todas as simulações são construídos após 10 repetições.

Relativamente à previsão UPA a figura 4.7 (a) mostra a comparação entre a série original e os resultados obtidos. Por sua vez, a figura 4.7 (b) revela os resultados para a previsão MPA, sendo o número de passos a prever de 10.

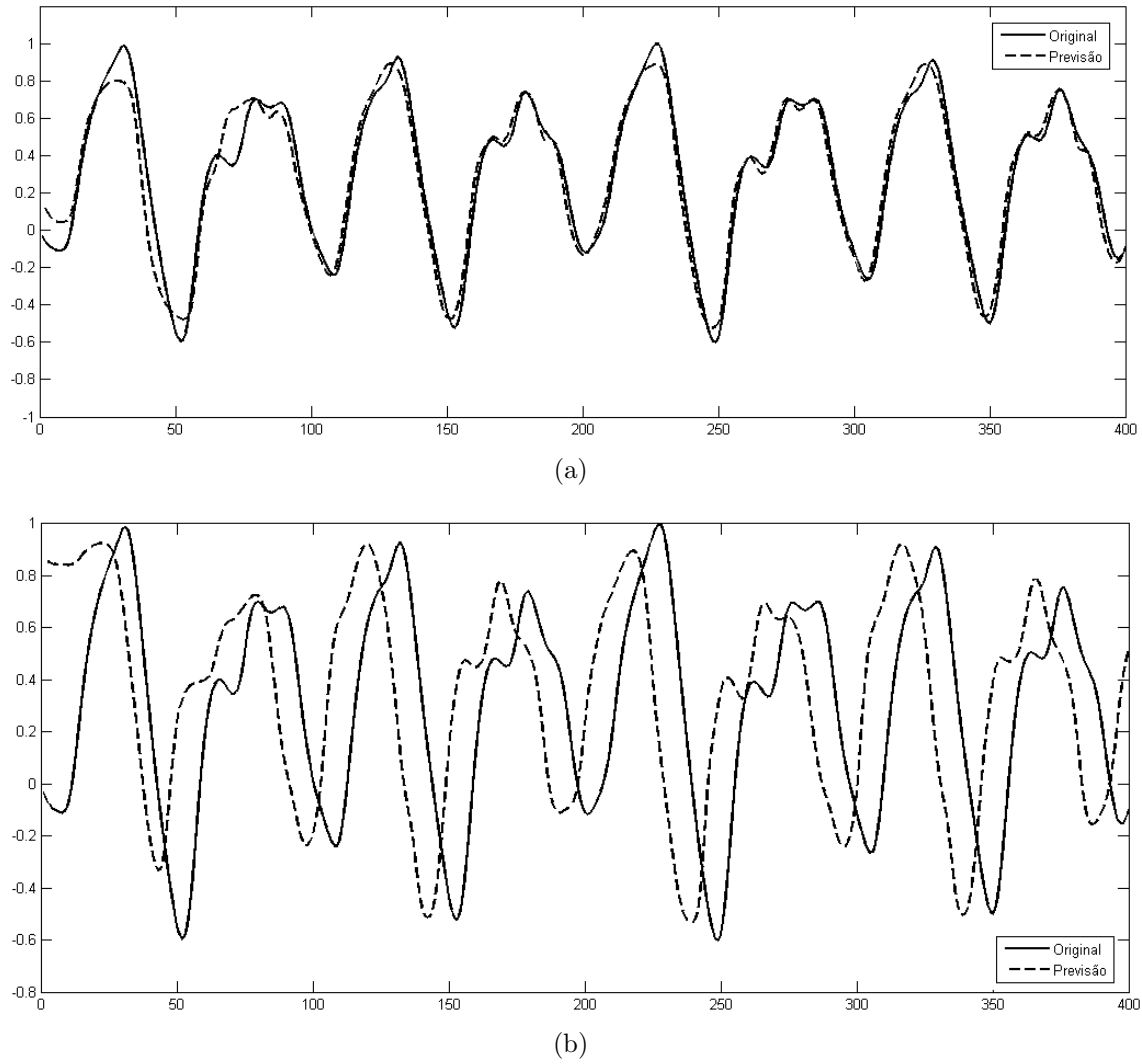


Fig. 4.7: Previsão da série de Mackey-Glass: (a) UPA ; (b) MPA.

Na tabela 4.1 encontram-se discriminados duas formas de cálculos de erro: médias da diferença entre a saída da rede com o sinal original e dos valores de NMSE obtidos no treino das previsões UPA e MPA. Para esta série verifica-se que na descoberta do valor seguinte da série a rede apresenta um erro de apenas $8,05 \pm 0,60\%$. No entanto, para a previsão de 10 passos à frente o erro revela-se maior sendo de $44,33 \pm 0,69\%$.

4.4.2 Resultados da previsão do laser

No próximo teste avalia-se a rede criada com a série do laser caótico, largamente usada em estudos de *benchmark*. Na figura 4.8 apresentam-se os resultados da previsão obtidos para a série do laser caótico. A linha sólida representa os valores exatos do conjunto de teste original e a linha tracejada representa os valores da amplitude estimados.

Os resultados na tabela 4.2 sugerem que a saída da FTDNN oscila intermiten-

Tab. 4.1: Erros na previsão da série de *Mackey-Glass*.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,0805 \pm 0,0068$	0,0812	$0,4004 \pm 0,0219$	0,3943
10	$0,4433 \pm 0,069$	0,446	$0,4366 \pm 0,0150$	0,4378

Tab. 4.2: Erros na previsão da série de laser caótico.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,2856 \pm 0,0556$	0,2586	$0,3740 \pm 0,0334$	0,3751
10	$0,6333 \pm 0,0649$	0,6196	$0,3209 \pm 0,0283$	0,3074

temente, isto é, especialmente na previsão MPA verifica-se que, devido às bruscas oscilações na série original, o resultado é comprometido comprovado pela parte inicial da série prevista.

4.4.3 Resultados da previsão do sinal CTG

A série de CTG foi aquela que mereceu um maior cuidado quer de pré-processamento como de obtenção de resultados de previsão. A escolha do número de atrasos induzidos à rede foi estipulada de acordo com o elevado número de dados integrantes da série. Adotou-se o valor de 100 generalizando este valor para as outras séries também em estudo e procedeu-se igualmente à previsão UPA e MPA, neste último caso com previsão de 5, 10 e 100 passos à frente.

Na primeira experiência a FTDNN tem 50 neurónios na única camada escondida e as figuras 4.9 e 4.10 mostram o resultado para os casos de previsão de 1 e 10 passos à frente da série. A diferença entre estas previsões não é muito significativa, comprovando-se pelo cálculo dos erros cometidos na tabela 4.3.

Tab. 4.3: Erro na previsão da série CTG.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,1722 \pm 0,0091$	0,1715	$0,1429 \pm 0,0070$	0,1439
5	$0,1736 \pm 0,0104$	0,1759	$0,1433 \pm 0,0122$	0,1438
10	$0,1779 \pm 0,0082$	0,1779	$0,1436 \pm 0,072$	0,1445
100	$0,2714 \pm 0,0133$	0,2685	$0,1380 \pm 0,0085$	0,1393

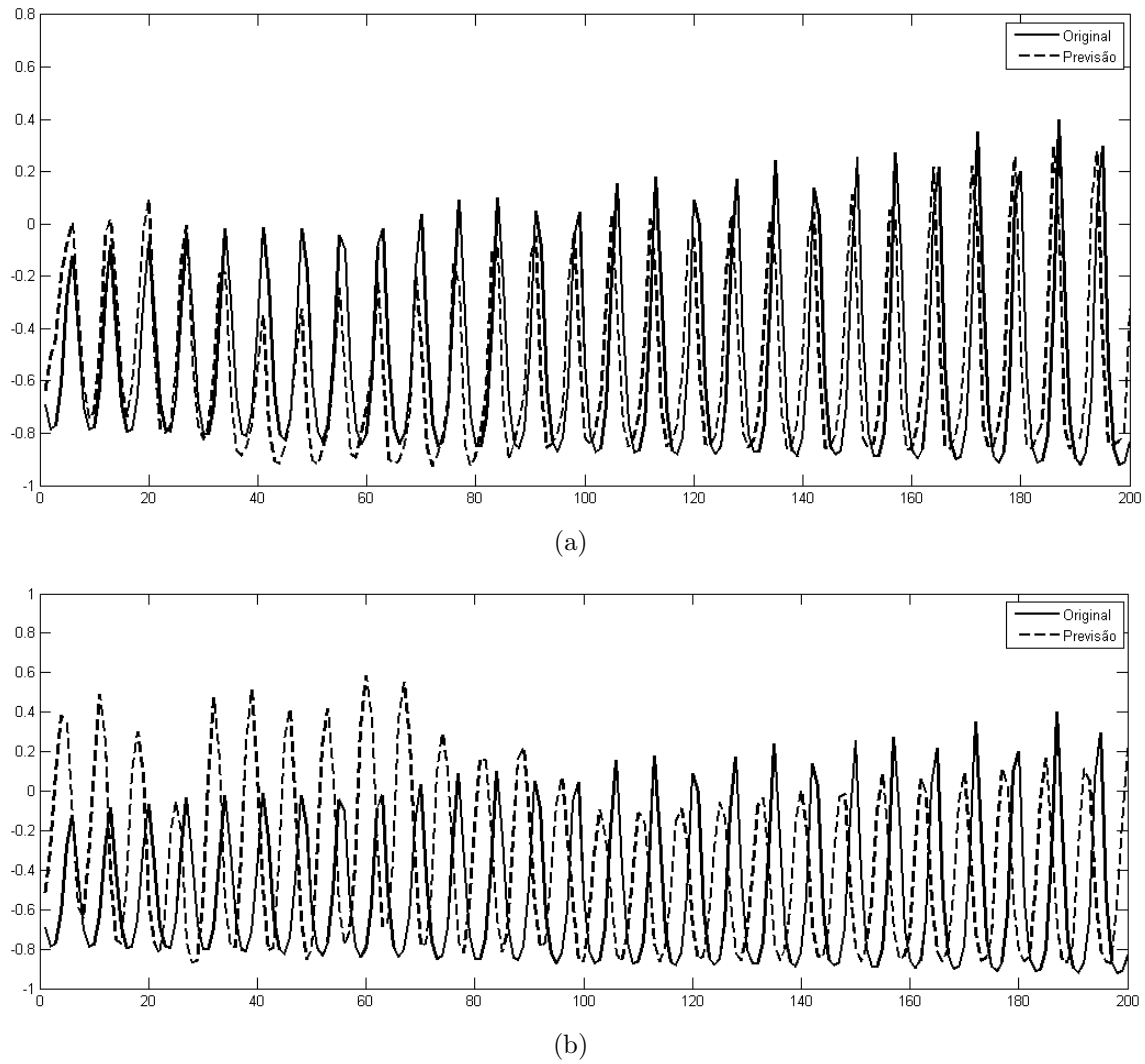


Fig. 4.8: Previsão da série do laser caótico: (a) UPA ; (b) MPA.

No que respeita à previsão de 100 passos à frente, figura 4.11, é ainda mais notória a diferença entre a série original e a obtida pela previsão da RNA, sendo o erro de previsão de $27,14 \pm 1,33\%$,

Com o objetivo de comparar se os resultados melhorariam, foi acrescentada à estrutura da rede mais uma camada escondida, admitindo metade do número de neurónios da camada escondida anterior. Desta forma a FTDNN tem 50 neurónios na primeira camada escondida e 25 na segunda. As figuras 4.12 e 4.13 revelam o resultado visual nos casos de previsão de 1 e 100 passos à frente da série.

Como se comprova pela tabela 4.4, os resultados das simulações de previsão com a camada adicional, admitindo os mesmos parâmetros de treino da rede, não se revelam melhores que a situação anterior. Embora a *performance* da rede no treino se tenha mostrado melhor, ou seja, o valor de NMSE é menor que os anteriores, o erro de previsão foi maior que na MLP de apenas uma camada escondida. A RNA teve uma maior capacidade de computação refletindo-se num melhor treino mas não

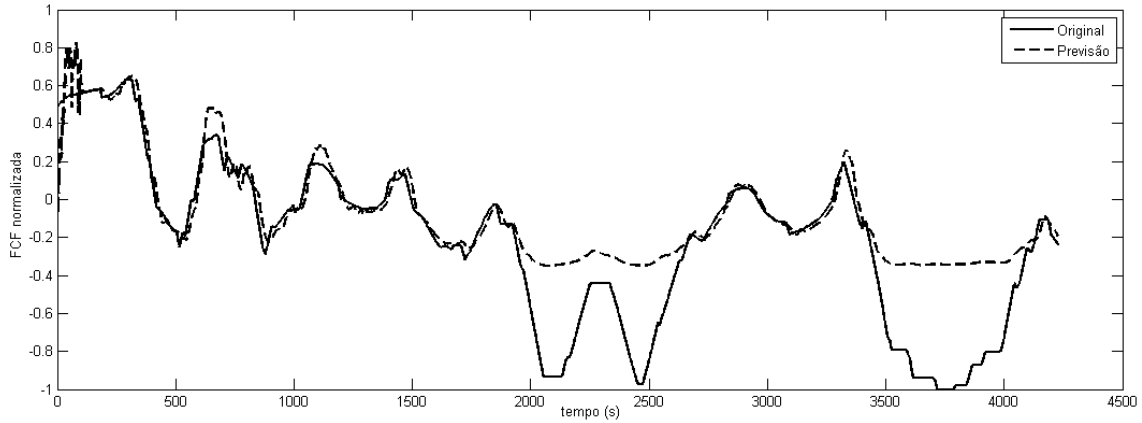


Fig. 4.9: Previsão da série CTG UPA.

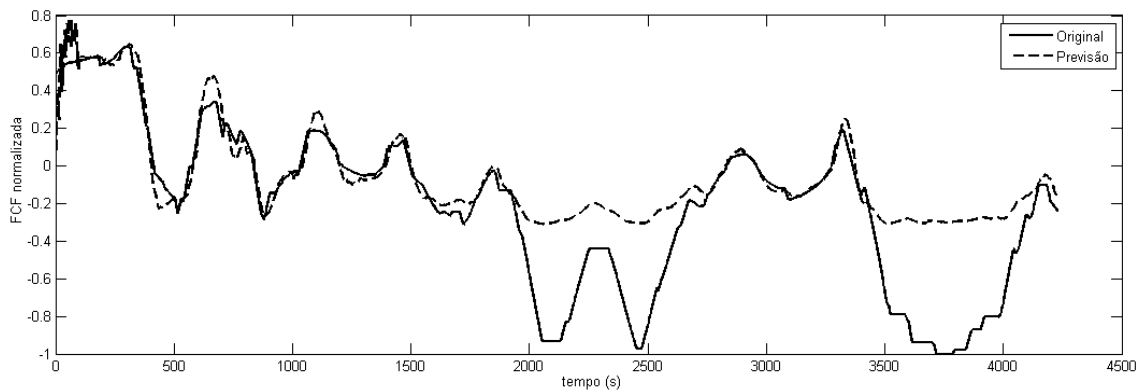


Fig. 4.10: Previsão da série CTG 10 passos à frente.

num menor erro.

Tab. 4.4: Erros na previsão da série CTG admitindo duas camadas escondidas na RNA.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,1897 \pm 0,0271$	0,1844	$0,1343 \pm 0,0194$	0,1355
5	$0,1804 \pm 0,0170$	0,1758	$0,1412 \pm 0,0186$	0,1469
10	$0,1873 \pm 0,0140$	0,1876	$0,1363 \pm 0,0150$	0,1362
100	$0,3188 \pm 0,1054$	0,3134	$0,2254 \pm 0,0807$	0,1959

As conclusões que se pretendem retirar não se podem ficar apenas pela aplicação do algoritmo a uma série de CTG. Desta feita, realizaram-se testes para uma outra série representada pela figura 4.14 que se comporta de uma forma mais homogênea comparativamente à anterior. Procedeu-se da mesma forma à suavização do sinal, estando este representado na figura 4.15. O tamanho da janela de suavização teve de ser ajustado para este exemplo adquirindo um valor de 20. Utilizaram-se os mesmos parâmetros anteriores na arquitetura e treino da RNA.

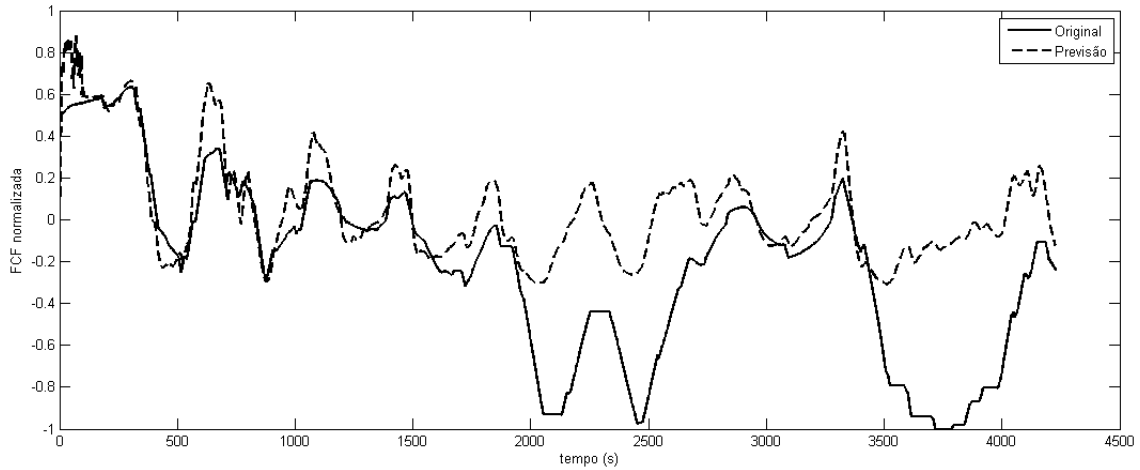


Fig. 4.11: Previsão da série CTG 100 passos à frente.

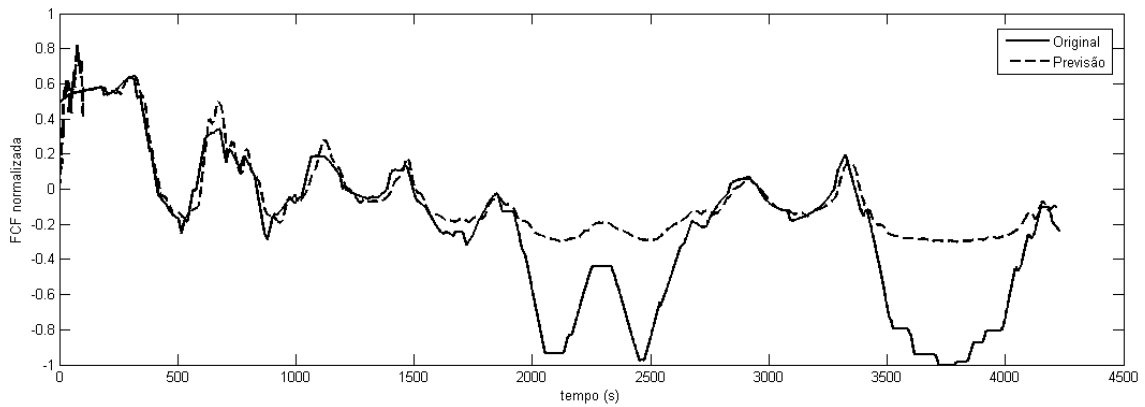


Fig. 4.12: Previsão da série CTG UPA com RNA de duas camadas escondidas.

Os resultados desta previsão para MLPs com uma e duas camadas escondidas, discriminados nas tabelas 4.5 e 4.6, revelam que a segunda série CTG prevista foi melhor aproximada do que a anterior quer seja na previsão de 1 ou 100 pontos. No caso de se admitir 2 camadas escondidas na estrutura da RNA a percentagem de erro resultante é igualmente menor.

Tab. 4.5: Erros na previsão da segunda série CTG.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,0269 \pm 0,0057$	0,0243	$1,5627 \pm 0,0354$	1,5618
100	$0,0607 \pm 0,0056$	0,0595	$1,5016 \pm 0,0230$	1,5025

Nas figuras 4.16 e 4.17 visualizam-se a série utilizada como modelo e a respetiva previsão obtida para os casos de UPA com a MLP de apenas uma camada intermédia e MLP com duas camadas escondidas na previsão de 10 pontos à frente, respetivamente. As aproximações efetuadas revelam-se mais precisas muito em parte devido

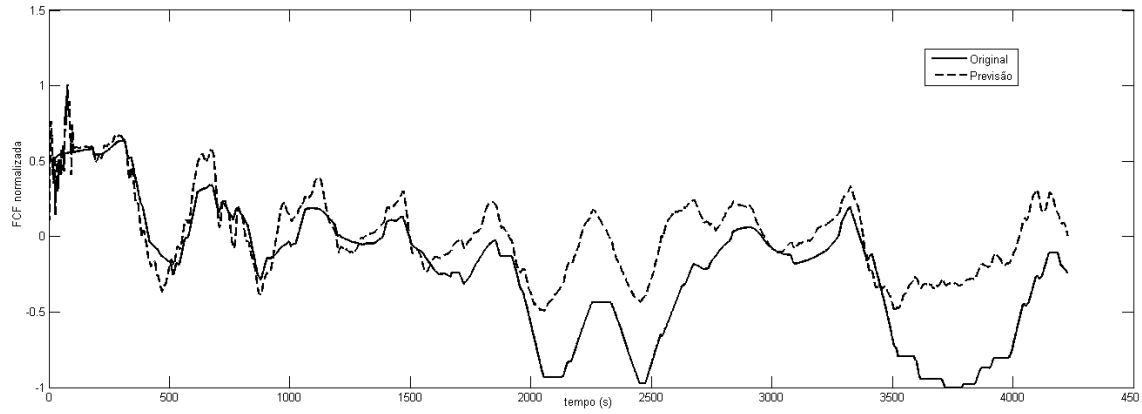


Fig. 4.13: Previsão da série CTG de 100 passos adiante com RNA de duas camadas escondidas.

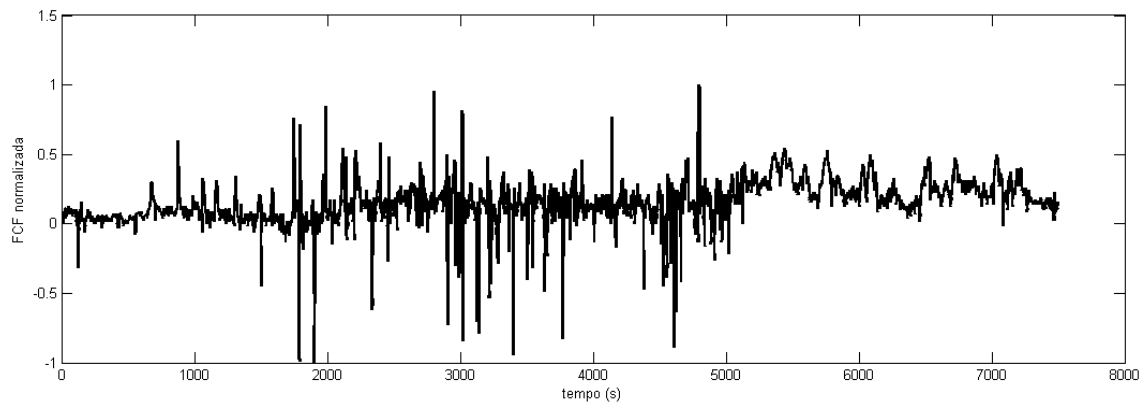


Fig. 4.14: Representação da série de FCF do CTG utilizada nesta experiência.

ao sinal ter valores mais homogêneos quando comparado com a outra série estudada.

4.5 Conclusões

Neste capítulo foram apresentados os resultados obtidos com o método de previsão proposto. A FTDNN faz parte de uma classe geral de redes cuja dinâmica aparece apenas na camada de entrada da MLP, ou seja, indução de um atraso à entrada da rede, sendo esta a escolha da arquitetura de rede para previsão. No entanto as RNA dinâmicas recorrentes alcançariam, segundo diferentes estudos efetuados, melhores resultados de previsão devido às suas características.

Não se optou pela utilização de redes como é o caso das NARX ou *Elman* porque exigia que o pré-processamento dos dados bem como escolha dos parâmetros fosse mais criterioso e conseqüentemente mais lento.

A série de CTG utilizada foi modificada de forma a poder aplicar-se melhor ao algoritmo. A utilização de diferentes métodos de suavização do sinal certamente influencia o treino da série, concluindo que se poderia ter obtido melhores resultados de previsão através do uso de diferentes formas de processamento do sinal.

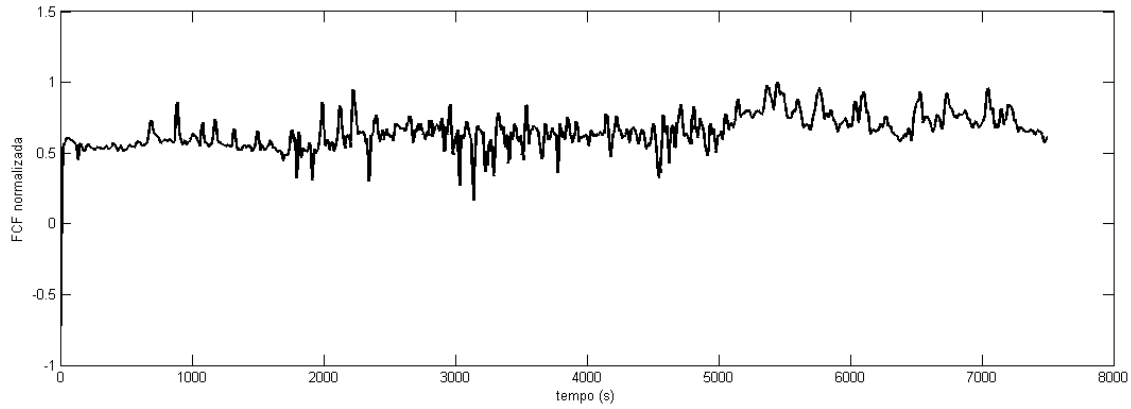


Fig. 4.15: Representação da série de FCF do CTG utilizada após filtragem.

Tab. 4.6: Erros na previsão da segunda série CTG com duas camadas escondidas na RNA.

Previsão	Erro		<i>NMSE</i>	
	Média	Mediana	Média	Mediana
1	$0,0316 \pm 0,0036$	0,0310	$1,5610 \pm 0,0452$	1,5504
100	$0,0610 \pm 0,0057$	0,0595	$1,5234 \pm 0,0175$	1,5214

Da mesma forma, a divisão dos dados para treino e teste é outro dos parâmetros que influenciam os resultados. Deduz-se que quanto maior a percentagem de treino da rede menor será o erro cometido logo melhor será a previsão final. Em muitas situações utilizam-se todos os dados disponíveis para efetuar o treino da rede e posteriormente faz-se o teste da rede com apenas uma parte da série.

Um dos problemas da previsão com RNA consiste na utilização na entrada de um conjunto de valores da série de tempos anteriores para prever os próximos valores. Muitas vezes o que acontece é que se tratam de séries irregulares em que a tentativa de previsão de valores nunca ocorridos no passado (durante o treino da rede) acaba por ser uma grande dificuldade. Esta situação verificou-se sobretudo na série de CTG.

Na análise da primeira série do sinal CTG pode verificar-se que na sua maioria o sinal é regular apenas apresenta valores menores na parte final do traçado. Torna-se complicado fazer previsões com magnitudes nunca atingidas nas situações de treino, o que se verifica pelos resultados obtidos. Por sua vez, na segunda experiência realizada com um outro sinal de CTG, os resultados revelaram-se mais satisfatórios tendo em conta a maior uniformidade do sinal.

No geral, o algoritmo de previsão utilizado faz uma boa aproximação quer das séries de referência amplamente utilizadas para testar o desempenho de modelos de RNA (*Mackey-Glass* e *laser*) como das principais séries deste capítulo, a FCF do sinal

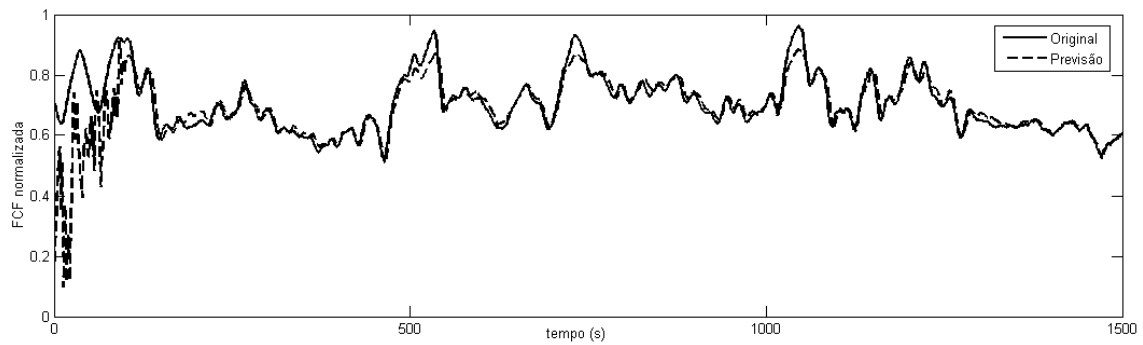


Fig. 4.16: Previsão da segunda série de CTG UPA.

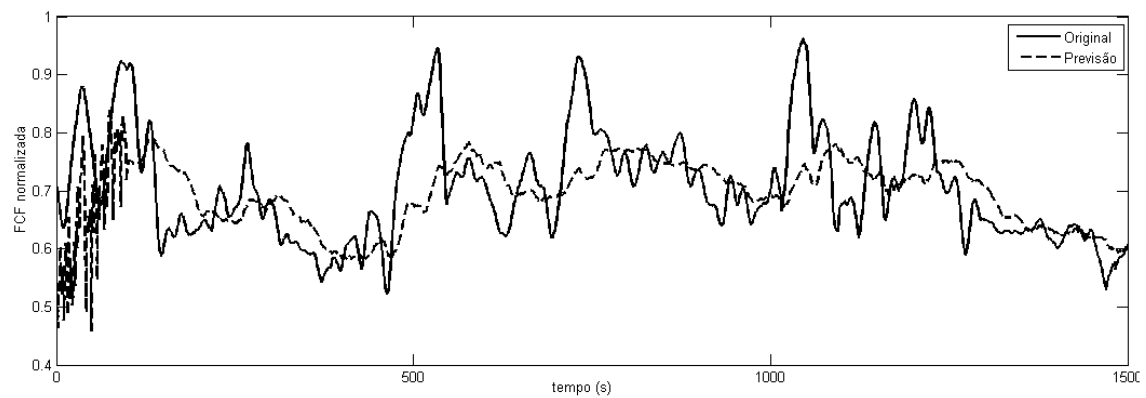


Fig. 4.17: Previsão da segunda série de CTG 10 instantes à frente admitindo duas camadas escondidas na RNA.

de CTG. Verifica-se que à medida que o horizonte de previsão aumenta as diferenças entre os valores previstos pela rede e os realmente observados se torna maior não se revelando um modelo muito robusto para previsão a longo prazo. Conclui-se, que, provavelmente, seria necessário optar-se por outros mecanismos que conseguíssem aprender melhor a dinâmica da série.

Conclusão

A importância deste estudo prende-se com o contributo do CTG como um exame complementar de diagnóstico de modo a possibilitar um diagnóstico diferencial de casos patológicos do feto cada vez mais corretos e precoces. A maior parte dos problemas reais de engenharia, informática e medicina são dinâmicos necessitando de ferramentas capazes de modelá-los eficazmente. Esta dissertação centra-se na classificação e previsão de dados CTG e utiliza como principal ferramenta as RNAs motivada pela sua capacidade de aproximação de funções, reconhecimento de padrões, extração de informação temporal e possuir melhor desempenho do que os procedimentos estatísticos convencionais.

A fraca fiabilidade da análise visual do CTG permanece uma grande preocupação e o aparecimento de sistemas computadorizados representam um novo desafio na avaliação do bem-estar fetal. O Omniview-SisPorto® descrito sucintamente no capítulo 2 revela-se uma tecnologia promissora, desenvolvida com o objetivo de sustentar a decisão clínica e de reduzir a complexidade da vigilância fetal.

Os resultados da classificação do CTG com FFNN evidenciam que a RNA teve, no geral, uma boa capacidade de discriminação de padrões. Efetuaram-se diferentes experiências de forma a testar qual a melhor arquitetura de rede bem como os melhores parâmetros a utilizar pois trata-se de um problema de classificação complexo, em que além do elevado número de dados é necessário ter-se em conta as diferentes classificações do sinal. A rede foi treinada com diferentes proporções dos dados para treino e teste sendo que obteve melhor resultado, isto é, menor erro de classificação, quando se considerou 70% dos dados para treino e os restantes para teste. As experiências preliminares com os dados *Íris* e *Yeast* foram uma mais valia por permitirem modelar o algoritmo para diferentes situações de classificação. O modelo de classificação pode ser melhorado recorrendo ao uso de mais variáveis explicativas que tragam benefícios em entender o CTG, de modo a que a RNA minimize o erro, servindo-se da capacidade de aprendizagem de um maior número de padrões.

A previsão de séries temporais é também uma área em que as RNAs têm demonstrado grande eficácia. No capítulo 4 delineou-se o desenvolvimento de um algoritmo para a aplicação em CTG recorrendo a redes supervisionadas dinâmicas não-recorrentes, que possuem atrasos na entrada para prever dinâmicas temporais. A aquisição do traçado CTG está suscetível a diferentes interferências, sendo por isso muito importante o seu pré-processamento antes da aplicação de técnicas de previsão. A falha do sinal bem como o ruído existente são fatores que podem comprometer os resultados.

O algoritmo de previsão utilizado apresenta resultados aceitáveis nas séries utilizadas. A divisão da série em diferentes conjuntos (treino e teste) influenciam os resultados obtidos na medida em que se verifica que quanto maior o número de dados utilizados no treino da rede melhor será a previsão efetuada. Comprova-se que quanto maior é o intervalo de tempo que se pretende prever, o erro da previsão aumenta, concluindo que a FTDNN não é muito robusta em previsão a longo prazo. Nesta parte, um outro fator que se revela influenciador dos resultados é a irregularidade do sinal CTG. Em muitas situações, o traçado CTG apenas é anormal na parte final, ou seja, todo o treino da rede é efectuado com valores mais ou menos homogêneos e, consequentemente, o erro da previsão de dados ainda não apresentados à rede será maior, como se verificou no primeiro exemplo.

O facto de se tentar efetuar previsão de um sinal fisiológico acresce particulares dificuldades por ser necessário ter-se em conta múltiplos fatores: desde o tempo de gestação, as condições clínicas maternas, alterações repentinas de condições fetais entre muitos outros. Ajudaria se existisse algum método que definisse, de acordo com os instantes que se pretendem prever, a janela temporal do sinal mais importante para essa previsão. Na prática estas situações não acontecem e não são aceites. A definição do estado do feto só é feita ao conjugarem-se as características antes e após o parto.

Toda a pesquisa científica apresenta dificuldades que não podem ser desconsideradas e omitidas. A aquisição de dados recentes de CTG devidamente classificados por um perito foi um dos problemas encontrados. A falta de bases de dados atuais fez com que se tivesse de recorrer a um conjunto de dados de CTG utilizados em trabalhos anteriores. Um outra grande dificuldades encontrada passou pela configuração das RNAs. Para definir corretamente a rede a ser utilizada, tanto na classificação como na previsão do CTG, foi necessário escolher o algoritmo de treino, número de camadas e respetivos números de neurónios, funções de transferência, tipos de pré-processamento dos dados como normalização, filtragem e divisão. O facto de ainda não existir uma metodologia capaz de apresentar a melhor configuração da rede levou a que este processo se tornasse iterativo o que o tornou extremamente moroso.

A aplicação destas técnicas de IA na prática clínica de rotina exige a realização de um considerável número de ensaios clínicos devidamente controlados e que requerem um aperfeiçoamento contínuo. Deve ter-se em conta a quantidade e qualidade dos dados, o tempo de computação necessário para a obtenção de resultados, o *software* específico e o nível técnico dos peritos envolvidos. Não se pretendeu demonstrar que as RNAs são a melhor ferramenta para efetuar estas tarefas pois não foi feita comparação com outros modelos. De acordo com os resultados obtidos, a *performance* foi positiva, tendo-se concluído que as RNAs podem contribuir para uma análise automática do sinal de CTG, tanto numa abordagem de classificação como numa abordagem de previsão a curto prazo.

5.1 Desenvolvimentos Futuros

Em primeiro lugar é importante referir que achamos que os principais objetivos deste trabalho foram atingidos tendo-se estudado com algum detalhe as principais etapas na utilização das RNAs para as tarefas de classificação e previsão dos CTGs. Há a noção de que existem muitos outros tópicos de pesquisa abertos e que merecem trabalhos futuros.

Para continuação do trabalho realizado nesta dissertação sugere-se que, no que respeita à classificação, exista um aumento da base de dados de CTG para permitir uma generalização dos resultados através de protocolos com clínicas ou hospitais e diferentes metodologias de *performance* para os classificadores.

Relativamente à previsão de CTG podem aplicar-se diferentes e mais eficazes métodos de pré-processamento do sinal de forma a obter melhor suavização, utilização de RNAs recorrentes para a previsão, como o caso das redes NARX (possivelmente obtenção de melhores resultados tendo em conta a arquitetura recorrente da rede) e uma comparação dos resultados obtidos com outros métodos de previsão.

Resultados Intermédios da Classificação

A.1 Resultados obtidos para escolha dos parâmetros de classificação

A.1.1 Seleção dos algoritmos de treino

Para se comparar diferentes algoritmos de treino compararam-se valores de erros de classificação para os dados Íris em diferentes experiências. Na tabela A.1 encontram-se os resultados dos erros e desvios padrões para 10 repetições do algoritmo, admitindo como funções de ativação a *logsig*, $\eta = 0,01$, tendo 4 neurónios na camada escondida da RNA.

Tab. A.1: Erros dos algoritmos de treino relativamente ao conjunto de dados Íris.

N	<i>trainlm</i>	<i>traingda</i>
	mean \pm std	mean \pm std
10	0,19 \pm 0,2137	0,1247 \pm 0,1595

A.1.2 Seleção dos valores da taxa de aprendizagem e número de neurónios

O objetivo desta etapa foi a seleção do melhor conjunto de parâmetros para se poder criar a MLP mais adequada para a classificação. Relativamente ao conjunto de dados Íris, os resultados de *performance* estão assinalados na tabela A.2, tendo sido apenas variado o número de neurónios, admitindo $\eta = 0,01$. A tabela A.3 mostra os resultados relativos ao conjunto de dados *Yeast*, tendo variado ambos os parâmetros, número de neurónios e taxa de aprendizagem.

Tab. A.2: Resultados de *performance* da rede (MSE) para o conjunto de dados *Íris*.

η	nº neurónios	traingda	traingd
0,01	4	0,1945	0,8564
	5	0,2768	0,1654
	10	0,2859	0,0970
	15	0,2854	0,1633

Tab. A.3: Resultados de *performance* da rede (MSE) para o conjunto de dados *Yeast*.

η	nº neurónios	traingda	traingd
0,01	10	0,1292	0,5077
	20	0,1153	0,1122
	30	0,0031	0,0100
	40	0,0260	0,1024
	50	0,0367	0,1170
0,05	10	0,2103	0,6108
	20	0,2043	0,1285
	30	0,006	0,0212
	40	0,0302	0,2132
	50	0,0394	0,1027

A.2 Experiências Preliminares da classificação com o algoritmo desenvolvido

No decorrer deste trabalho diferentes experiências foram efetuadas para se verificar que conjunto de parâmetros e funções obteriam melhores resultados na classificação do CTG. Os resultados são de situações em que a MLP criada apresenta a arquitetura [16:10], com os dados normalizados para o intervalo $[0,1]$ e $[-1,1]$ e podem consultar-se nas tabelas, A.4, A.5 e A.6 os respetivos resultados da classificação para as diferentes experiências. Estas experiências não complementaram a seleção do melhor número de neurónios para as camadas escondida e da taxa de aprendizagem, tendo sido estes valores escolhidos por tentativa-erro.

Tab. A.4: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 60% para treino e 40% para teste.

N	FT	<i>traingd</i>	<i>traingda</i>
		mean \pm std	mean \pm std
10	<i>logsig</i>	0,5776 \pm 0,1656	0,1937 \pm 0,0302
	<i>tansig</i>	0,2936 \pm 0,0505	0,2792 \pm 0,0370
20	<i>logsig</i>	0,5689 \pm 0,1498	0,1981 \pm 0,0298
	<i>tansig</i>	0,2962 \pm 0,0576	0,2854 \pm 0,0483

Tab. A.5: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 70% para treino e 30% para teste.

N	FT	<i>traingd</i>	<i>traingda</i>
		mean \pm std	mean \pm std
10	<i>logsig</i>	0,5609 \pm 0,1236	0,1890 \pm 0,0286
	<i>tansig</i>	0,2985 \pm 0,0624	0,3032 \pm 0,0747
20	<i>logsig</i>	0,5759 \pm 0,1599	0,1937 \pm 0,0334
	<i>tansig</i>	0,2944 \pm 0,0592	0,2915 \pm 0,0651

Tab. A.6: Erros (média \pm desvio padrão) de classificação relativos ao conjunto de dados CTG com 90% para treino e 10% para teste.

N	FT	<i>traingd</i>	<i>traingda</i>
		mean \pm std	mean \pm std
10	<i>logsig</i>	0,5583 \pm 0,1595	0,1888 \pm 0,0284
	<i>tansig</i>	0,2931 \pm 0,0500	0,2795 \pm 0,0308
20	<i>logsig</i>	0,5602 \pm 0,1446	0,1879 \pm 0,0289
	<i>tansig</i>	0,3073 \pm 0,0775	0,2813 \pm 0,0494

Apêndice B

Resultados Intermédios da Previsão

B.1 Divisão das séries temporais para previsão

Nas figuras B.1 e B.2 encontram-se representadas as séries temporais de *Mackey-Glass* e do laser com a respetiva divisão em conjunto de treino e teste.

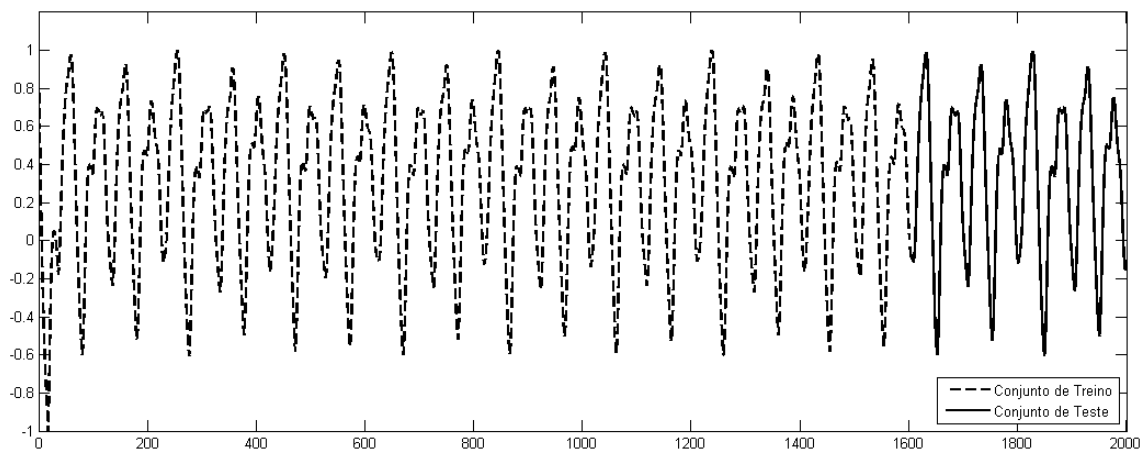


Fig. B.1: Série de *Mackey-Glass* com distinção dos conjuntos de treino e teste.

B.2 Outros Resultados de Previsão obtidos

Nesta secção são apresentados diferentes resultados obtidos na previsão das séries de CTG. A figura B.3 mostra o resultado da previsão de 5 passos adiante da primeira série de CTG estudada.

Do mesmo modo que se realizou o pré-processamento da primeira série CTG, a figura B.4 apresenta diferentes resultados de suavização do segundo sinal. Seguidamente, a figura B.5 revela a previsão de 50 passos à frente relativa a esta série. Efetuaram-se outras tentativas utilizando desta vez a série temporal relativa às UC do sinal de CTG, estando nas figuras B.6 e B.7 resultados elementares da aplica-

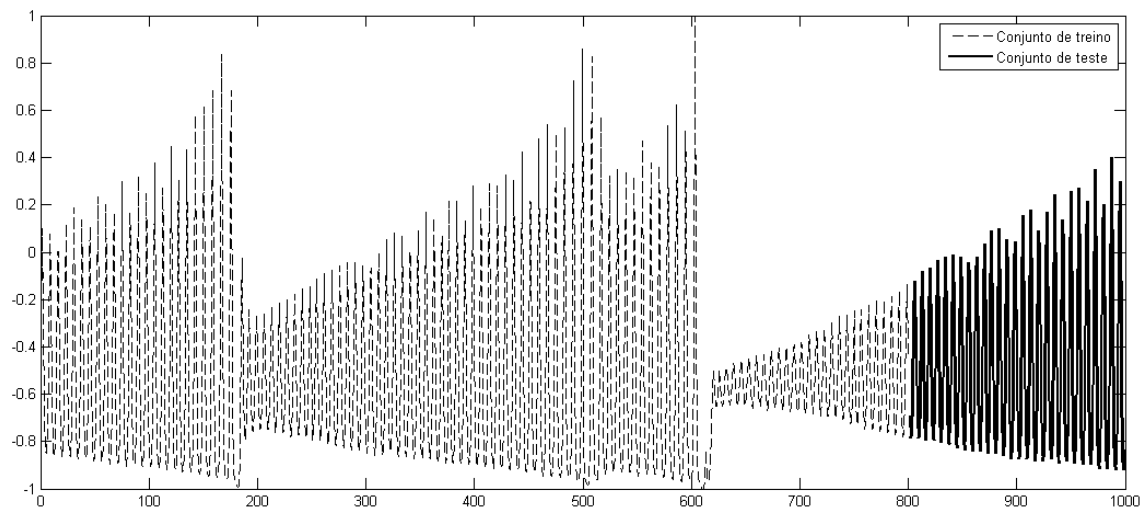


Fig. B.2: Série do laser com distinção dos conjuntos de treino e teste.

ção do algoritmo de previsão. No entanto, este tópico não foi incluído no estudo apresentado.

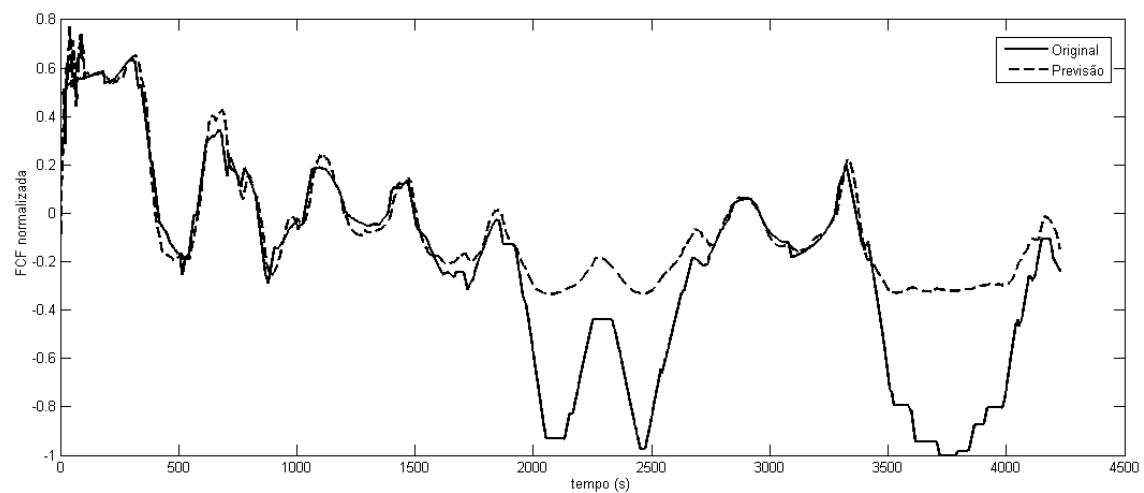
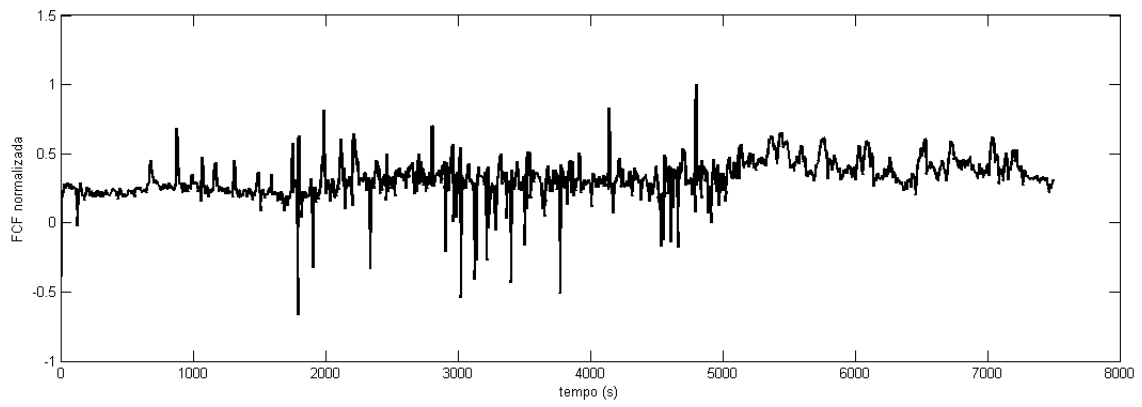
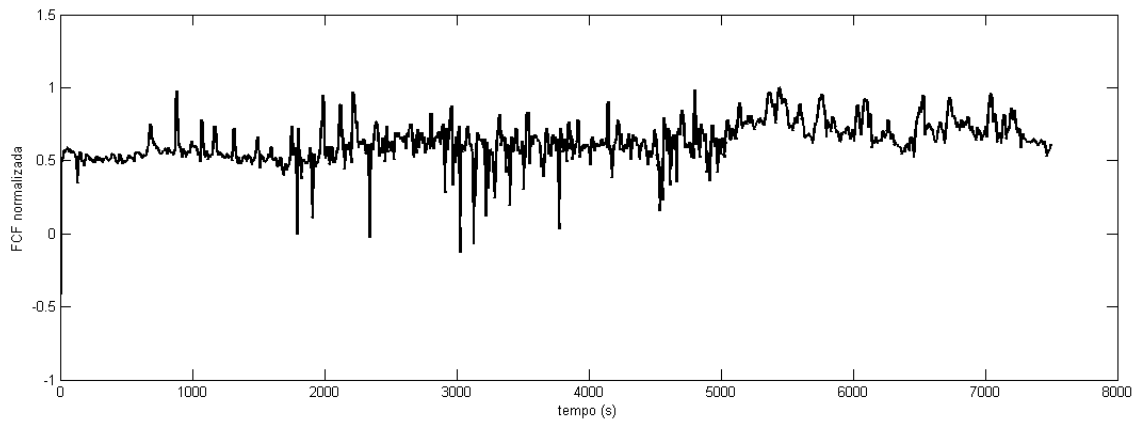


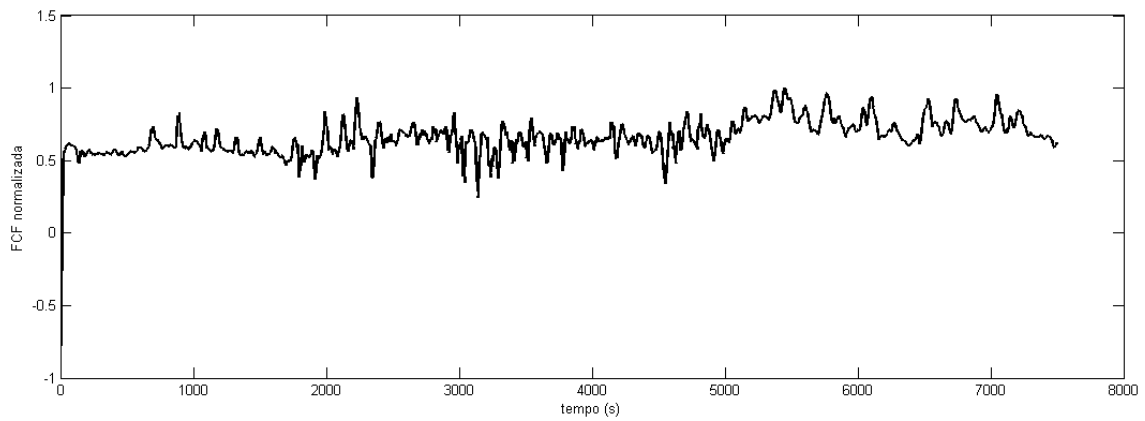
Fig. B.3: Previsão da série de CTG 5 passos à frente.



(a)



(b)



(c)

Fig. B.4: Resultados da suavização do segundo sinal de FCF do CTG: (a) $M=5$;(b) $M=10$; (c) $M=25$.

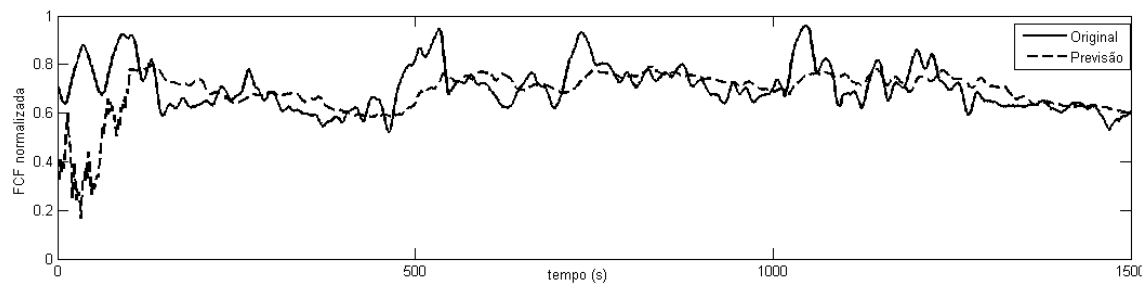


Fig. B.5: Previsão da FCF do segundo sinal de CTG 50 passos à frente.

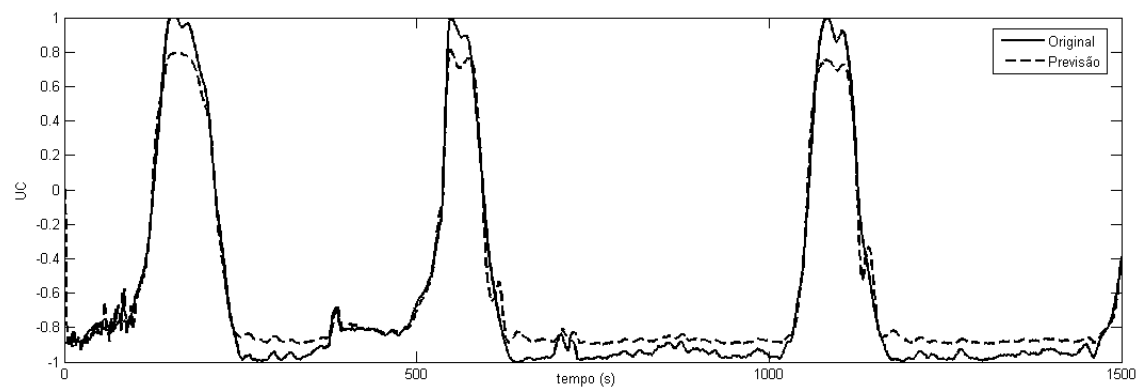


Fig. B.6: Previsão das UC da segunda série CTG UPA.

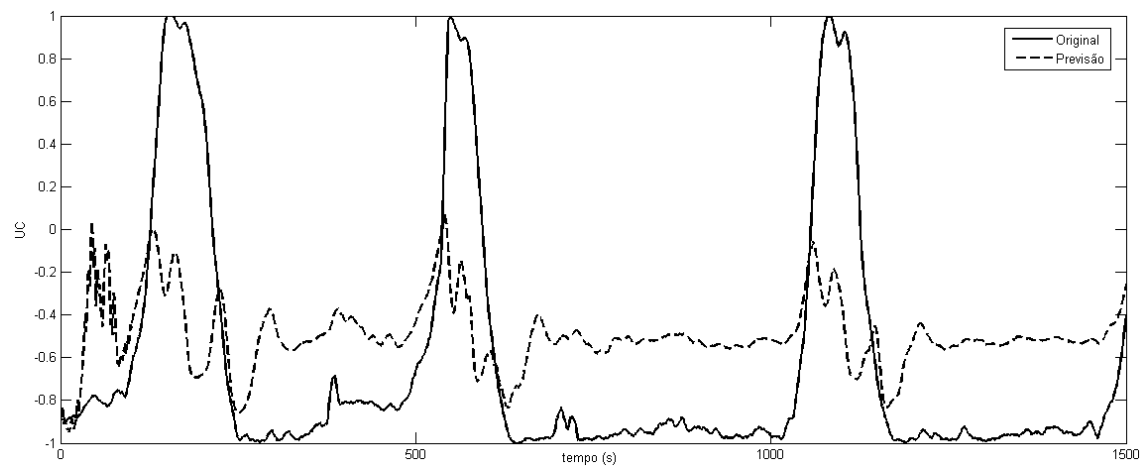


Fig. B.7: Previsão das UC da segunda série CTG 50 passos à frente.

Bibliografia

- [1] J. Bernardes, *Leitura Automatizada do Cardiotocograma- Contribuição para o seu desenvolvimento e avaliação*. PhD thesis, Faculdade de Medicina da Universidade do Porto, 1993.
- [2] A. Costa, *Development and Evaluation of a Combination of Computer Analysis of Cardiotocography and Electrocardiography for Intrapartum Fetal Monitoring*. PhD thesis, Faculdade de Medicina da Universidade do Porto, 2010.
- [3] R. Czabanski, J. Jezewski, A. Motonia, and M. Jezewski, “Computerized analysis of fetal heart rate signals as the predictor of neonatal acidemia,” *Expert Systems with Applications*, 2012.
- [4] L. M. Graça, “Monitorização fetal intra-parto,” in *Medicina Materno-Fetal* (4^o, ed.), ch. cap. 31, Lidel, 2010.
- [5] A. C. Souza, “Cardiotocografia estimulada em gestações de baixo risco: estudo comparativo da resposta cardíaca fetal à estimulação vibratória e sônica,” Master’s thesis, Universidade de São Paulo, 2006.
- [6] D. A. de Campos and J. Bernardes, “Twenty-five years after the figo guidelines for the use of fetal monitoring: Time for a simplified approach?,” *International Journal of Gynecology and Obstetrics*, 2010. Elsevier.
- [7] J. Bernardes, H. Gonçalves, D. A. de Campos, and A. P. Rocha, “Linear and complex heart rate dynamics vary with sex in relation to fetal behavioural states,” *Early Human Development*, no. 84, pp. 433–439, 2008.
- [8] A. Costa, D. A. de Campos, F. Costa, C. Santos, and J. Bernardes, “Prediction of neonatal acidemia by computer analysis of fetal heart rate and st event signals,” *Am J Obstet Gynecol*, vol. 201, pp. 464e1–6, 2009.
- [9] D. A. de Campos and J. Bernardes, “Comparison of fetal heart rate baseline estimation by sisporto1 2.01 and a consensus of clinicians,” *European Journal of Obstetrics and Gynecology and Reproductive Biology*, no. 117, pp. 174–178, 2004. Elsevier.
- [10] D. A. de Campos, C. Costa-Santos, and J. B. F. the SisPorto Multicentre Validation Study Group), “Prediction of neonatal state by computer analysis of

- fetal heart rate tracings: the antepartum arm of the sisporto multicentre validation study,” *European Journal of Obstetrics and Gynecology and Reproductive Biology*, no. 118, pp. 52–60, 2005.
- [11] D. A. de Campos, P. Sousa, A. Costa, and J. Bernardes, “Omniview-sisporto® 3.5-a central fetal monitoring station with online alerts based on computerized cardiotocogram + st event analysis,” *J. Perinat. Med.*, no. 36, pp. 260–4, 2008.
 - [12] T. Machado, “Modelação de séries temporais- métodos lineares e não lineares,” Master’s thesis, Instituto Politécnico de Bragança, 2009.
 - [13] P. M. Rodrigues, “Diagnóstico da doença de alzheimer com base no electroencefalograma,” Master’s thesis, Instituto Politécnico de Bragança, Julho 2011.
 - [14] G. Georgoulas, D. Gavrilis, I. G. Tsoulos, C. Stylios, J. Bernardes, and P. P. Groumpos, “Novel approach for fetal heart rate classification introducing grammatical evolution,” *Biomedical Signal Processing and Control*, no. 2, pp. 69–79, 2007.
 - [15] R. Frank, N. Davey, and S. Hunt, “Time series prediction and neural networks.” Department of Computer Science, University of Hertfordshire, Hatfield, UK.
 - [16] J. M. Júnior, “Redes neurais dinâmicas para predição e modelagem não-linear de séries temporais,” Master’s thesis, Universidade Federal do Ceará, 2006.
 - [17] J. A. Sousa, “Aplicação de redes neuronais na previsão de vendas para retalho,” Master’s thesis, Faculdade de Engenharia da Universidade do Porto, 2011.
 - [18] S. Cazares, L. Tarassenko, L. Impey, M. Moulden, and C. W. G. Redman, “Automated identification of abnormal cardiotocograms using neural networks visualization techniques,” *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001.
 - [19] M. Jezewski, J. Wrobel, P. Labaj, J. Leski, N. Henzel, K. Horoba, and J. Jezewski, “Some practical remarks on neural networks approach to fetal cardiotocograms classification,” *Procee*, 2007.
 - [20] G. Magenes, M. G. Signorini, and D. Arduini, “Classification of cardiotocographic records by neural networks,” *IEEE*, 2000.
 - [21] C. Ulbricht, G. Dorffner, and A. Lee, “Neural networks for recognizing patterns in cardiotocograms,” *Artificial Intelligence in Medicine*, vol. 12, no. 3, pp. 271–84, 1998.
 - [22] S. Nidhal, M. Ali, A. Zaidan, B. Zaidan, and H. Najah, “Computerized algorithm for fetal heart rate baseline and baseline variability estimation based on distance between signal average and a value,” *International Journal of Pharmacology*, vol. 7 (2), pp. 228–237, 2011.
 - [23] K. Ojala, *Modern Methods in the Prevention and Management of Complications in Labor*. PhD thesis, University of Oulu Finland, 2010.

- [24] B. G. Berdinas, A. A. Betanzos, and O. F. Romero, "Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system," *Artificial Intelligence*, no. 136, pp. 1–27, 2002.
- [25] Speculum, "Artigos médicos s.a.." www.speculum.pt, 2012. Consultado em Abril de 2012.
- [26] N. Medical, "Neoventa® st analysis." <http://www.neoventa.com/products/st-analysis/>, 2012. Consultado em Agosto de 2012.
- [27] A. C. of Obstetricians and Gynecologists, "<http://www.acog.org/>," 2012. Consultado em Agosto de 2012.
- [28] D. A. de Campos, J. Bernardes, and A. Costa, "Análise e interpretação da cardiotocografia," tech. rep., Unidade de Obstetrícia Serviço de Ginecologia e Obstetrícia Hospital de S. João, 2005.
- [29] R. C. of Obstetricians and Gynaecologists, *The use of electronic fetal monitoring Evidence-based clinical guideline*. National Institute for Clinical Excellence, 8 ed., 2001.
- [30] Omniview-SisPorto®, "Fetal central monitoring station." www.omniview.eu/, 2012. Consultado em Abril de 2012.
- [31] P. Moura, "Vigilância fetal intraparto," *Arquivos Maternidade Dr. Alfredo da Costa*, vol. XVI, p. 72, Dezembro 2005.
- [32] "Okb medical ltd." <http://www.okbmedical.com/omniview-sisporto.html>, 2010. Consultado em Junho de 2012.
- [33] T. S. Project, "Improving fetal heart rate analysis and interpretation." <http://sisporto.med.up.pt/>. Consultado em Abril de 2012.
- [34] S. Haykin, *Neural Networks A comprehensive Foundation*. Prentice Hall, 2 ed., 1999.
- [35] P. Rodrigues, "Redes neuronais aplicadas à segmentação e classificação de leucócitos em imagens," Master's thesis, Universidade do Porto, 2000. <http://hdl.handle.net/10198/696> Consultado em Abril de 2012.
- [36] J. Santos, *Data Classification with Neural Networks and Entropic Criteria*. PhD thesis, Faculdade de Engenharia da Universidade do Porto, Janeiro 2007.
- [37] D. Semedo, "Credit scoring: Aplicação da regressão logística vs redes neuronais artificiais na avaliação do risco de crédito no mercado cabo-verdiano," Master's thesis, Universidade Nova de Lisboa, 2009.
- [38] A. B. Geva, "Scalenet- multisclae neural-network architecture for time series prediction," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, 1998.
- [39] J. M. Júnior and G. A. Barreto, "Long-term time series prediction with the narx network: An empirical evaluation," *Elsevier Science*, 2007.

- [40] Z. Tang and P. A. Fishwick, "Feed-forward neural nets as models for time series forecasting." TR91-008 Computer and Information Sciences, University of Florida.
- [41] A. Braga, *Curvas ROC: Aspectos Funcionais e Aplicações*. PhD thesis, Universidade do Minho, Dezembro 2000.
- [42] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23 rd International Conference on Machine Learning*, Pittsburgh, PA, 2006.
- [43] A. M. Simundic, "Measures of diagnostic accuracy: basic definitions," tech. rep., University Department of Chemistry, Sestre milosrdnice University Hospital, Zagreb, Croatia.
- [44] M. Jezewski, R. Czabanski, J. Wrobel, and K. Horoba, "Analysis of extracted cardiotocographic signal features to improve automated prediction of fetal outcome," *Biocybernetics and Biomedical Engineering*, vol. 30, no. 4, pp. 29–47, 2010.
- [45] Y. Noguchi, F. Matsumoto, K. Maeda, and T. Nagasawa, "Neural network analysis and evaluation of the fetal heart rate," Tech. Rep. 19-30, Algorithms, 2, 2009.
- [46] T. Masters, *Practical Neural Network Recipes in C++*. Academic Press, 1993.
- [47] L. H. B. Bertolucci, "Comparação entre redes neurais e técnicas clássicas na previsão a curtíssimo prazo de demanda de potência eléctrica ativa," tech. rep., Departamento de Engenharia Elétrica SEL EESC USP, 2008.
- [48] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, no. 10, pp. 215–236, 1996.
- [49] Mathworks, "Matlab the language of technical computing." <http://www.mathworks.com/products/matlab/>, 2012.
- [50] M. H. Beale, M. T. Hagan, and H. B. Demuth, "Neural network toolbox getting started guide," tech. rep., MathWorks, 2012.
- [51] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annual Eugenics*, vol. 7, no. II, pp. 179–188, 1936.
- [52] A. Frank and A. Asuncion, "UCI machine learning repository." <http://archive.ics.uci.edu/ml>, 2010. University of California, Irvine, School of Information and Computer Sciences.
- [53] J. Peralta, G. Gutierrez, and A. Sanchis, "Time series forecasting by evolving "shuffle", cross-validation and ensembles," in *Artificial neural networks - ICANN 2010 : 20th international conference* (L. 6352, ed.), pp. 50–53, Springer, 2010.
- [54] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," 2008. Arizona State Univeristy.

- [55] T. Ayer, J. Chhatwal, O. Alagoz, J. Charles E. Kahn, R. W. Woods, and E. S. Burnside, "Comparison of logistic regression and artificial neural network models in breast cancer risk estimation," *radiographics.rsna.org*, vol. 30, pp. 13–22, Janeiro 2010. Published online November 9, 2009,.
- [56] J. M. Pereira, M. Domínguez, and J. Ocejó, "Modelos de previsão do fracasso empresarial: Aspectos a considerar," *Revista de Estudos Politécnicos*, vol. IV, no. 7, pp. 111–148, 2007.
- [57] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, pp. 44–55, February 2001.
- [58] S. Samarasinghe, *Neural Networks for Applied Science and Engineering: from Fundamentals to Complex Pattern Recognition*. Auerbach Publications, 2007.
- [59] P. Fernandes, J. Teixeira, J. M. Ferreira, and S. G. Azevedo, "Modelling tourism demand: A comparative study between artificial neural networks and the box-jenkins methodology," in *Romanian Journal of Economic Forecasting*, Institute of Economic Forecasting, 2008.
- [60] P. O. Fernandes and J. P. Teixeira, "New approach of the ann methodology for forecasting time series: Use of time index." Instituto Politécnico de Bragança.
- [61] T. Machado, J. P. Teixeira, and P. Fernandes, "Modelação da procura turística em portugal: Regressão linear versus redes neuronais artificiais." Instituto Politécnico de Bragança.
- [62] J. P. Teixeira and P. Fernandes, "Nova abordagem da metodologia de redes neuronais artificiais para a previsão de séries temporais de turismo: a data como índice. aplicação à região da madeira," *16º Congresso da APDR*, 2010.
- [63] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *IEEE Trans. on Power Systems*, vol. 17, pp. 626–632, 2002.
- [64] M. A. Shamisi, A. H. Assi, and H. Hejase, *Engineering Education and Research Using MATLAB*, ch. 9 Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in AI Ain City- UAE. InTech, 2011.
- [65] J. C. Lam, K. Wan, and L. Yang, "Solar radiation modelling using anns for different climates in china," *Energy Conversion and Management*, no. 49, pp. 1080–1090, 2008.
- [66] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, "Times series prediction with multilayer perceptron, fir and elman neural networks,"
- [67] R. Lina, L. Yanxin, R. Zhiyuan, L. Haiyan, and F. Ruicheng, "Application of elman neural network and matlab to load forecasting," *International Conference on Information Technology and Computer Science*, 2009.

- [68] R. H. Loh, "Time series forecast with neural network and wavelet techniques," *The University of Queensland*, 2003.
- [69] A. da Silva Soares, "Predição de séries temporais econômicas por meio de redes neurais artificiais e transformada wavelet: Combinando modelo técnico e fundamentalista," Master's thesis, Universidade de São Paulo, 2008.
- [70] R. Yadav, P. Kalra, and J. John, "Time series prediction with single multiplicative neuron model," *Applied Soft Computing*, no. 7, pp. 1157–1163, 2007.
- [71] R. Drossu and Z. Obradovic, "Rapid design of neural networks for time series prediction," 1996. IEEE Computational Science and Engineering.
- [72] X. Wua and Y. Wang, "Extended and unscented kalman filtering based feedforward neural networks for time series prediction," *Applied Mathematical Modelling*, no. 36, pp. 1123–1131, 2012.
- [73] R. Zemouri, D. Racocanu, and N. Zerhouni, "Recurrent radial basis function network for time series prediction," *Engineering Applications of Artificial Intelligence*, no. 16, pp. 453–463, 2003.
- [74] E. J. Strini, "Previsão da incidência de dengue por meio de redes neurais artificiais," tech. rep., Universidade de São Paulo, 2006.
- [75] E. M. Forney and C. W. Anderson, eds., *Classification of EEG During Imagined Mental Tasks by Forecasting with Elman Recurrent Neural Networks*, IEEE, Proceedings of International Joint Conference on Neural Networks, 2011.
- [76] M. C. Mackey and L. Glass, "Oscillations and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
- [77] M. Ardalani-Farsa and S. Zolfaghari, "Chaotic time series prediction with residual analysis method using hybrid elman-narx neural networks," *Neurocomputing*, no. 73, pp. 2540–2553, 2010.
- [78] Z. Huang and M.-L. Shyu, "k-nn based ls-svm framework for long-term time series prediction." University of Miami, USA.
- [79] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural Network Toolbox: User's Guide R2011b*. Mathworks, 2011.
- [80] V. Chudacck, J. Spilka, B. Rubackova, and M. Koucky, "Evaluation of feature subsets for classification of cardiocographic recordings," *Computers in Cardiology*, no. 35, pp. 845–848, 2008.
- [81] H. Gonçalves, J. Bernardes, A. P. Rocha, and D. A. de Campos, "Linear and nonlinear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period," *Early Human Development*, no. 83, pp. 585–591, 2007.
- [82] H. Gonçalves, A. P. Rocha, D. A. de Campos, and J. Bernardes, "Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery," *Med Bio Eng Comput*, vol. 44, no. 847-855, 2006.

-
- [83] J. M. Júnior and G. A. Barreto, “Multistep-ahead prediction of rainfall precipitation using the narx network,” 2008. Finland.
- [84] J. M. Júnior and G. A. Barreto, “Extensões da rede recorrente de elman para predição não-linear de séries temporais caóticas: um estudo comparativo,” *10ª Conferência Brasileira de Dinâmica, Controle e Aplicações*, 2011. DIN-CON2011.